

A COMPARATIVE STUDY OF NINE GROUP
TESTS OF INTELLIGENCE FOR
PRIMARY GRADES

BY
MARY LORETTE DOUGHERTY

A DISSERTATION

submitted to the Board of University Studies of The Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

FEBRUARY, 1928

BALTIMORE
THE JOHNS HOPKINS PRESS
1928

A COMPARATIVE STUDY OF NINE GROUP
TESTS OF INTELLIGENCE FOR
PRIMARY GRADES



Digitized by the Internet Archive
in 2026

A COMPARATIVE STUDY OF NINE GROUP
TESTS OF INTELLIGENCE FOR
PRIMARY GRADES

BY
MARY LORETTE DOUGHERTY

A DISSERTATION

submitted to the Board of University Studies of The Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

FEBRUARY, 1928

BALTIMORE
THE JOHNS HOPKINS PRESS
1928

COPYRIGHT 1928 BY
THE JOHNS HOPKINS PRESS
BALTIMORE, MD.



The Lord Baltimore Press
BALTIMORE, MD., U. S. A.

APPRECIATION

In the inception, prosecution and completion of this study I am indebted to many who have assisted from time to time.

First, I want to express my great indebtedness to Professor Edward F. Buchner. His sympathetic interest in the planning of the study and his arrangement with the Department of Public Schools of Baltimore, Maryland, for the opportunity to carry on the experimental work, made possible the organization of the study as undertaken. His consideration in arrangement of my program of teaching so as to make possible my prosecution of this experiment has been a great factor in the success of the work. The advice gained from time to time from conferences in which the work was discussed and the expression of his faith in the value of the study has been an inspiration as well as a practical advantage in the progress of the experiment. My gratitude for his co-operation can be expressed only feebly by a statement of indebtedness.

Professor Florence E. Bamberger has from the first given valuable advice as to methods both in the experiment and in the handling of the data. Her preview of the field of work has helped in the formulation of the lines of investigation to be pursued. During Dr. Buchner's year of absence she directed the work and has from the first given freely of time and effort in consultation, reading of manuscript and helpful criticism.

Dr. Fowler D. Brooks has advised in all matters of statistical treatment of the data, both in the choice of methods and in the matter of technique. His assistance has been of great value in the organization and manipulation of the data.

Aside from the unlimited assistance I have received from the members of the Department of Education, I have received courtesies and help from others to whom I express my appreciation. This is due to the Administrative Officers

of the Baltimore Public Schools for permission to conduct the experiment; and to the following members of the Baltimore Bureau of Statistics and Research: To Mr. E. B. Hebden for furnishing the distribution tables on which my choice of children was based, to Miss Isabel Lazarus for furnishing the scores from one test which had been given shortly before I began the testing and for other co-operation. I am also indebted to the principals and teachers of the five schools which participated in the experiment for their cheerful and hearty co-operation. I should like also to express my appreciation of the co-operation of the 800 children who participated in the first year's work and especially to the nearly 400 who repeated the work the second year. Their attitude toward the work speaks well for the influence of the schools and made the work of giving the tests a pleasure rather than a task.

I am indebted also to a number of persons who through the last three years have given clerical assistance in connection with checking and recording the data and its results.

M. L. D.

TABLE OF CONTENTS

	PAGE
Introduction	I
Chapter 1. Problem and Materials.....	3
Chapter 2. Technique of the Experiment.....	11
Chapter 3. Criteria Used in the Study.....	16
Chapter 4. The First Year's Study.....	21
Chapter 5. The Second Year's Study.....	76
Chapter 6. Conclusions	95
Bibliography	110

INTRODUCTION

In 1905 Alfred Binet presented a new departure in methods of studying the development of mental life, giving to the world the first attempt to measure degree of mental ability by observing and evaluating an individual's reaction to a definite complex situation. Since that time there has been considerable development and improvement of this method of study. The World War, with the necessity for rating large numbers of men in a short time, stimulated the organization and development of group tests of intelligence. The presence of numbers of men who were unable to read English led to the organization of non-language tests. It was from this phase of the test development that group tests of intelligence for use in the primary grades had their origin.

For several years following the war there was a somewhat prolific production in such tests, nine of which persisted and were on the market in 1924. There had been made no comparative study of these tests which might form a basis for intelligent choice among them. It was in this situation that the problem of the present study originated.

CHAPTER 1

PROBLEM AND MATERIALS

Early in 1924 the writer found in Pintner's "Intelligence Testing," then a new publication, a list of group tests of intelligence for kindergarten-primary grades,¹ 11 in number. The question presented itself as to whether there was any basis for intelligent choice among these tests or whether one might as well pick at random. An investigation to discover studies made which would answer this question was not very fruitful and the idea arose of undertaking such a study. The belief that an answer found to this question would be of practical value to primary school administrators and teachers encouraged a further investigation of the situation.

After conferences with Dr. Buchner and Dr. Bamberger correspondence was opened with publishers of all these tests preliminary to planning the experimental study. One publisher stated that his test had been withdrawn from the market because of lack of funds for needed revision. Another publisher responded that he did not recommend the test for the primary grades. This left nine tests for possible use as follows:

TEST	PUBLISHER
Cole-Vincent Test for School Entrants.	Kansas State Teachers College, Emporia, Kansas.
Dearborn Group Tests of Intelligence, Series I.	J. B. Lippincott, Philadelphia.
Detroit First Grade Test.	World Book Co., Yonkers-On-Hudson, N. Y.
Haggerty Intelligence Test, Delta I.	World Book Co., Yonkers-On-Hudson, N. Y.
Otis Primary Intelligence Test.	World Book Co., Yonkers-On-Hudson, N. Y.
Pintner-Cunningham Primary Mental Test.	World Book Co., Yonkers-On-Hudson, N. Y.
Kingsbury Primary Group Intelligence Scale.	Public School Publishing Co., Bloomington, Ill.
Pressey Primary Classification Test.	Public School Publishing Co., Bloomington, Ill.
Myers Mental Measure.	Newson and Co., New York.

¹ Pintner, R. *Intelligence Testing*. New York, H. Holt, 1923, p. 167.

Since all of these tests had been studied and to some degree validated in the process of preparation and were all in use with the assumption of their reliability, it was not regarded as a necessary preliminary to criticize their validity. The Manuals published for each test furnish such description and such corroborative data as the authors thought it necessary to publish.¹

Other reports of the use of these tests either preliminary to their appearance in their present form or since then have appeared from time to time.²

None of these studies had been made specifically for the purpose of comparing the various tests to provide a satisfactory relative evaluation of the tests.

¹ Manuals published as follows:

Cole-Vincent Test for School Entrants, Emporia, Kansas, State Teachers College.

Dearborn, W. F., The Dearborn Group Tests of Intelligence, Series I, Manual. Philadelphia, Pa. Lippincott.

Detroit First Grade Intelligence Test, Manual and Directions for Use. Yonkers-On-Hudson, N. Y., World Book Co.

Haggerty Intelligence Examination, Delta I. Manual, Yonkers-On-Hudson, N. Y., World Book Co.

Kingsbury, F., Primary Group Intelligence Scale, Bloomington, Ill., Public School Publishing Co.

Myers, G. C., Measuring Minds, Carlisle, Pa., The Sentinel.

Otis, A. S., Manual for Primary Group Intelligence Test, Yonkers-On-Hudson, N. Y. World Book Co.

Pintner, R., and Cunningham, B. V., Primary Mental Test, Manual, Yonkers-On-Hudson, N. Y., World Book Co.

Pressey, S. L., Primary Classification Scale, Bloomington, Ill. Public School Publishing Co.

² Myers, C. E. and G. C., A Group Intelligence Test, Sch. and Soc. 1919, 10, 355-360.

Pintner, R., and Cunningham, B. V., The Problem of Group Intelligence Tests for Very Young Children, Jour. Educ. Psychol. 1922, 13, 465-472.

Pintner, R., The Pintner-Cunningham Primary Test, Jour. Educ. Psychol. 1927, 18, 52-58.

Pressey, S. L. and L. W., Group Point Scale for Measuring General Intelligence with First Results from 1100 School Children, Jour. Appl. Psychol. 1918, 2, 250-269.

Pressey, L. W., Group Scale of Intelligence for Use in First Three Grades, Jour. Educ. Psychol. 1919, 10, 297-308.

Chassell, C. F. and L. M., Survey of First Three Grades of Horace Mann School, Jour. Educ. Psychol. 1921, 12, 72-81, 243-252.

Myers, G. C., Teachers vs. Mental Tests as Prophets of School Progress, Sch. and Soc. 1922, 16, 300-303.

Pressey, S. L. and L. W., Measuring the Usefulness of Tests in Solving School Problems, Sch. and Soc. 1920, 12, 531-534.

Three studies have been made which appear to have had the definite purpose of discovering the relative value of the tests.

Miss Margaret V. Cobb, at that time Secretary of the Bureau of Mental Tests and Measurements of the University of Michigan, in the spring of 1921 made such a study in Jackson, Mich., using the following tests: Dearborn Group Tests of Intelligence, Series I, Part I, Haggerty Intelligence Test, Delta I, Kingsbury Primary Group Intelligence Scale, Otis Primary Intelligence Test, and Detroit First Grade Intelligence Test. This study is referred to as "an elaborate comparative study,"¹ with a footnote stating "The results of this study are to appear in a doctorate thesis by Miss Cobb." An effort to trace this work ended finally with a letter from Miss Cobb stating that "it will be a long time before this is published," and giving the additional information, noted above, that only Part I of the Dearborn Test was used.

Miss Ruth Streitz,² as the basis for her master's thesis at the University of Chicago in 1922, made a study using the Pressey Primer Scale, Myers Mental Measure, Dearborn Group Intelligence Tests, Series I, and the Haggerty Intelligence Test, Delta I. The Pressey Primer Scale is an earlier edition of the Primary Classification Test which was used in the present study and which Mrs. Pressey regards as having "marked improvements" over the earlier form.³ The Haggerty also may be an earlier edition than the one used in this study as it is described as "composed of 10 tests and 10 pre-tests or fore-exercises." However the discussion states that in scoring credit is given for "rights" in four tests and for "rights minus wrongs" in two tests. This coincides with the arrangement of the test used in the present study which is composed of six sub-tests with a pre-test for each.

¹ Twenty-first Yearbook, National Society for the Study of Education, Part II, p. 139.

² Streitz, Ruth, *An Experimental Investigation of Four Group Intelligence Scales for Primary Children*. Master's Thesis, September, 1922, University of Chicago.

³ Pressey, L. W., *The Primary Classification Test*. Jour. Educ. Research. 1924, 9, 314.

For the Dearborn test an earlier form divided into three sections and allowing a total credit of 300 was given in the Madison schools in November. This was not available in February for the testing in the University of Chicago school and the revised form having two sections with a total score of 202 was used. This revised form is the one used in the present study.

Miss Streitz's subjects were 196 children located as follows :

School	First grade	Second grade	Total
University of Chicago	48	48	96
Madison Public Schools, 2 schools	50	50	100
Totals	98	98	196

In Madison two schools were used to provide a "comprehensive sampling." The children of one of the schools were mainly from homes of foreign parentage and having noticeable language difficulty. The children in the other school were from one of the best residential districts, from homes of good social standing, with parents interested in the school and in the children's progress. The University of Chicago school gave an even more select group with much attention to individual instruction.

Miss Streitz's purpose is to show the reliability and validity of the tests as a means of measuring mental ability in the primary grades. Her criteria are: 1. The correlation with teachers' rating of ability taken before the testing and made in 10 percentile groups from 60 to 99; 2. The intercorrelation of the tests; 3. The variation of the median score from first to second grade; 4. The correlation between various pairs of tests and the teachers' estimates—a. by the method of weighting,¹ b. by the method of averages.

A comparison is also made of the median scores on each test for the two grades showing the gain from first grade

¹ The Scales were weighted for combination so as to give approximately 100 as the median value of each, as follows:

	First grade	Second grade
Dearborn	multiplied by 2	and by $1\frac{1}{3}$
Myers	multiplied by 3	and by 2
Pressey	multiplied by 2	and by $1\frac{1}{4}$
Haggerty	multiplied by $1\frac{1}{4}$	and by 1

to second grade and also the percent of overlapping between the grades. The percent of overlapping is found for each school.

Assuming that the value of the test is in inverse relation to the percent of overlapping, Miss Streitz finds evidence that Dearborn and Haggerty tests are best.

The present study cannot be directly compared with Miss Streitz's for a number of reasons. Among these are the facts: 1. That two of the tests used were earlier editions than those used in the present study, and therefore, presumably less satisfactory; 2. That Miss Streitz used the raw scores on the tests as the basis of comparison while the present study uses the Group IQ; 3. That the external criterion used by Miss Streitz is the teachers' estimates of ability. The one used in the present study is the Stanford-Binet IQ.

The range of correlations in the various tables of Miss Streitz's report is relatively small and it is quite possible that such an amount of variation might have little significance. The study is an interesting beginning in the investigation of the relative value of the tests used but it gives no convincing evidence of any considerable superiority of one test over another. The method of overlapping is a valid method of showing the worth of the scales for classification if we assume that the children tested were at the time properly classified. One interesting result shown is that even though the data vary comparatively little, there is apparent consistency in showing the Dearborn and Haggerty tests to be the more desirable. The teachers' estimate of intelligence, while commonly accepted has frequently been criticized as to reliability. In the present study such estimates were secured before the testing and it is planned to make use of them in a following study. For the study reported here the Stanford-Binet IQ was chosen as the chief external criterion because of its generally accepted reliability for groups.

Miss Ada B. Viele, in Roanoke Rapids, N. C., in the spring of 1924, used four tests and reported the results¹ of her study

¹ Viele, Ada B., *A Study of Four Primary Mental Tests*, Ele. Sch. Jour. 1925, 25, 675-681.

which was for the purpose of discovering "the relative usefulness of several of the better known primary mental tests." The tests used were the Cole-Vincent, Dearborn, Detroit, and Pintner-Cunningham. She employed two criteria. 1. Teachers' estimates of pupils' ability to learn first or second grade subjects. The children were rated on ability numerically, Number 1, being the best and no two being given the same rating. The teachers were instructed "in case two or more seem to you to be equally intelligent, flip a coin to decide which will be given the higher rating." There was a second rating handled in the same way, on *effort*. 2. Inter-correlation of the tests.

Each teacher gave the tests to her own pupils and marked the papers. The teachers were given no special training in methods of administering these tests but all were given general instructions about following the printed directions carefully. This was probably not so unscientific as it appears because the teachers who did the work were members of an extension class in educational measurements (presumably from the University of North Carolina).

The subjects were located in schools as follows :

Grade	No. of classes	No. of pupils	Highest reported for one test
1	8	273	168
2	4	141	162
Opportunity	1	18 } 159	

Apparently not all the children were present for testing as the numbers given with the different tests do not agree with the totals given. There is no explanation of the larger number reported as tested in the second grade than are given as the enrollment.

The Cole-Vincent, Detroit and Pintner-Cunningham tests were given to every class. The Dearborn test was given to five of the eight first grade classes (160 children) and to two of the four second grade classes (92 children). The range of the correlations of the test scores with teachers' estimates of ability are given and also the limits of these ranges. The author believes that "the wide range in the correlations indicates either the complete failure of the tests specified to

disclose academic ability, or the inability of a few teachers, even after a year's acquaintance with them, to estimate accurately their pupils' mental ability." She prefers the latter as the more reasonable conclusion. It is at least interesting that the highest correlations in the first grade are all from the same class while the lowest correlations are each from a different class. In the second grade the highest correlations come from three different classes, those from the Cole-Vincent and Dearborn tests coming from the same class. The lowest correlations are from three of the four classes listed as second grade and from the opportunity class which is apparently grouped with the second grade.

The mean correlations of these tests with the teachers' rating of ability is given. According to this criterion the tests are less satisfactory for the first grade than for the second grade. The same relation exists in the correlation with effort.

The intercorrelations of the tests for each grade are uniformly higher in the first grade than in the second. This, if taken as a measure of excellence of the tests, belies the correlations with teachers' estimates which show higher correlations in the second grade. A possible conclusion is that while the correlated tests in first grade measure the same characteristics, they do not agree so well in measuring what the teachers' estimates measure.

Little comparison of this study with that of Miss Streitz is possible because the Dearborn test is the only one used in both studies. Table 1 presents a comparison of the results of correlating the Dearborn test with teachers' estimates of ability in the two studies.

TABLE 1
MEAN CORRELATION OF DEARBORN TEST SCORES WITH TEACHERS' ESTIMATES OF ABILITY

Grade	Streitz U. of C.		Madison ¹		Viele	
	r	Pupils	r	Pupils	r	Pupils
1.	.605	48	.660	50	.421	160
2.	.515	48	.600	50	.513	92

¹ An earlier edition of the test was used in Madison. This may account for the higher correlation.

Miss Viele's study is like Miss Streitz's in using raw scores for the basis of comparison. It adds to Miss Streitz's study by the use of a larger number of children and by the study of other tests. It is open to question as to its scientific value because of the fact that the tests were given and scored by several different individuals, who were relatively untrained. While there are some correlations high enough to be encouraging, the mean correlations both with ability and with effort are not convincing.

The present study differs with this one as with the first mentioned in the criteria used. It also includes two tests not used in either study, the Otis Primary Intelligence Test and the Kingsbury Primary Group Intelligence Scale, and extends the range of testing from kindergarten through the third grade. There are also included in the first testing more children for each grade than are included in either of the two studies reported. A further important advance is the repetition of the tests after an interval of a year to check the results. Other extensions of the method of handling the data will appear in the report of the present study.

CHAPTER 2

TECHNIQUE OF THE EXPERIMENT

The plan for this experiment called for :

1. The testing of 200 children from each of the school grades: kindergarten, first, second, and third—a total of 800 children.
2. The retesting of one-half of these children in each grade one year later.
3. The giving of the Stanford-Binet individual intelligence test to 50 children from each grade, who were given all the tests for their group each year.

It was the desire to secure as far as possible a random sample of the children of the grades used from the children of those grades in the white schools of the city of Baltimore.

For this purpose the choice of schools to be used was considered first. Dr. Bamberger kindly gave her advice in selecting among the schools offered by the principals for the experiment, in such a way as to get a variety, racially and socially, in the schools used. The racial character of the school was determined by the birthplace of the father as recorded at the school. One of the schools had a population largely Polish and Bohemian with a sprinkling of other nationalities. One was very largely made up of Italian and Russian children. The other three schools were predominantly American, but of differing social status—one being a boulevard school, one from a mercantile neighborhood, and one from near the wharf and factory districts.

In order to secure the right distribution of children according to chronological age from the different grades, Mr. E. B. Hebden, then of the Department of Research and Statistics of the Baltimore schools, was consulted. He kindly furnished for use the tables of distribution of the children in these grades, according to chronological age, in the white schools of the city. The percentage of children of each age for each of the grades to be used was found from these tables and the distribution was determined on this basis. There was necessarily some adjustment as, for instance, where the per-

centage of children of a particular age was so small as to make less than a single child in the smaller group which was to be used. Table 2 will show the basis of choice and the use made of it. The first choice conformed to these requirements. Some variations occurred through loss of children by removal from the school where they were being tested.

The first column of Table 2 gives the chronological ages found in the grade, the second column, the total number of children of each grade and the grand total. The third column, showing percentage at age, was found in the following manner. Under the direction of Dr. Brooks the small numbers at the extremes were neglected because the percent was too small to be considered in the sampling used which was 200 children out of 1962, 9352, 8525, and 9326, from the kindergarten, first, second, and third grades, respectively. The total of the abridged group was found and the percents of each of the constituents in this total. In the kindergarten it was necessary to abridge further because this study was made at the beginning of the year when the six year olds had just been removed by transfer to the first grade. Therefore, in this grade the four and five year old groups were used. The other three grades were handled in the same way except that, because of the overlapping of chronological ages it was not necessary to eliminate any age group having as many as 3.2% of the total number of the children in the grade.

It was then determined how many children in each 50 in each grade should be of a given chronological age. This is shown in the table in the right hand column.

Table 3 shows the composition of the complete group, for the first year, which is used in this study. In Part I is given the distribution of chronological age, for each grade of this group as to number and percentage, and in Part II are given the percentages which it was planned to use. If the exception in the kindergarten, which was noted, is considered there is a rather good correspondence between the distributions.

TABLE 2

NET ROLL OF PUPILS OF ALL AGES IN KINDERGARTEN, AND FIRST
SECOND AND THIRD GRADES; BALTIMORE CITY SCHOOLS, 1922-23.

Age	Total	Percent at age	Percent used	For each 50 children
KINDERGARTEN				
3	I			
4	158	1307 {	8.053	12.088 — 12
5	1149		58.562	87.912 — 88
6	634		32.314	6 four year olds.
7	19			44 five year olds.
8	I			
	1962			
FIRST GRADE				
5	33	8930 {	0.353	
6	3725		39.831	41.713 — 42
7	4145		44.322	46.416 — 46
8	1060		11.334	11.870 — 12
9	295		3.154	
10	62		0.663	
11	18			
12	6			
13	4			
14	2			
15	2			
	9352			
SECOND GRADE				
6	106	8221 {	1.243	
7	2783		32.645	33.973 — 34
8	3627		42.545	44.118 — 44
9	1385		16.246	16.847 — 17
10	426		4.997	5.182 — 5
11	130	8327	1.525	
12	41			
13	15			
14	10			
15	2			
	8525			
THIRD GRADE				
6	6	9158 {	3.978	4.040 — 4
7	371		29.509	30.050 — 30
8	2752		37.435	38.119 — 38
9	3491		17.370	17.755 — 18
10	1626		6.787	6.912 — 7
11	633		3.056	3.112 — 3
12	285			
13	110			
14	39			
15	13			
	9326			

TABLE 3
DISTRIBUTION OF CHRONOLOGICAL AGES IN THE COMPLETE GROUPS FOR
THE FIRST YEAR

Grade age	Kindergarten		First		Second		Third	
	No.	%	No.	%	No.	%	No.	%
4	40	23						
5	130	74						
6	5	3	81	43	1	.3		
7			83	45	91	39	1	.5
8			18	10	99	42	49	24
9			2	1	33	14	79	39
10			1	.5	9	4	49	24
11			0	0	2	.7	20	10
12			1	.5			5	2
13							0	0
14							1	.5
Totals	175		186		235		204	

Grade age	PART II Percent Needed for Each Grade			
	Kindergarten %	First %	Second %	Third %
4	12			
5	88			
6		42		
7		46	34	4
8		12	44	30
9			17	38
10			5	18
11				7
12				3

The determination of which of the nine tests to use in each grade was made by using for each grade all tests for which norms or standards were furnished for the grade. In accordance with this plan the following program for use of the tests was planned:

Tests	Kindergarten	First	Second	Third grade
Cole-Vincent	x	x		
Dearborn	x	x	x	x
Detroit		x		
Haggerty		x	x	x
Kingsbury		x	x	x
Myers		x	x	x
Otis		x	x	x
Pintner-Cunningham	x	x	x	
Pressey		x	x	

After the children for the experiment had been selected the entire number in each grade was divided into two groups as nearly equal as possible. Where there were two classes in a school one of these was placed in each group. These were designated Group I and Group II. In Group I the tests were given in alphabetical order. In Group II in kindergarten the Cole-Vincent test was given first and the order of the other two reversed. In the first grade Cole-Vincent was given first as in kindergarten because of the requirement that it shall be used very early in the year. Then in first grade as in second and third grades the Kingsbury test was given followed by the other tests in reverse alphabetical order. This arrangement was made so as to allow the checking of practice effect on the tests if it should seem desirable to do so.

As a rule, each school was visited one day each week and an effort was made to distribute the testing among the classes so as to have rather regular intervals between the tests. This was not possible in all cases due to certain holidays and to variation in the size and number of the classes in different schools. In no case did the entire group test program for a class extend more than three months and usually it was not over two months.

On the completion of the group-testing program, 50 children from each grade who had taken all the tests given in the grade were given the Stanford Revision of the Binet-Simon test. These were chosen as far as possible proportionately from each chronological age group. Because of the loss of some of these children in the second year's testing and the necessity of replacing them there is not a definite known correspondence between the two groups.

CHAPTER 3

CRITERIA USED IN THE STUDY

At the beginning of the first year's study each teacher was asked to turn in her estimate of the intelligence of her children as judged by their ability to learn. This was to be stated numerically, 1 being the highest rank and no two having the same rank. These estimates are not to be used in this report.

The external criterion which is used in this report is the Stanford-Binet IQ. This is designated Criterion I (CI). Since this test was given to only 50 children from each grade the use of this criterion necessarily limits the group studied to approximately one-fourth of those tested the first year and to approximately one-half of those tested the second year. The justification of the use of this limited number as representative of the whole group will be presented later.

The criterion chosen for use with the whole group is the average of the Group IQ's of all the tests given to the group; three tests in the kindergarten; nine, in the first grade; seven, in the second grade; and five, in the third grade. This is designated Criterion II (CII).

The question arose with regard to the smaller groups, 50 in each grade, as to whether it would be possible to find a combination of these two criteria which would be more satisfactory than either or which would give added value to the study if used in addition to these two. To test this, Criteria III, IV, and V were determined and a comparison made of these with the first two. These were obtained by the following method:

For grades 1-3. Criterion III is the average of Criterion I (Stanford-Binet IQ) and Criterion II (average Group IQ). This involves the assumption that the Stanford-Binet test is equivalent to nine group tests in the first grade, to seven group tests in the second grade, to five group tests in the third grade. Because of the common notion that group tests

are more valid for older children than for younger children it was felt that this was not an extremely rash assumption. Criterion IV is the average of Criterion I taken twice and Criterion II taken once. Criterion V is the average of Criterion I taken once and Criterion II taken twice. That is, with Criterion III a middle point was taken between Criterion I and Criterion II, with Criterion IV this point was thrown to the extreme toward Criterion I and with Criterion V this point was thrown to the extreme toward Criterion II. The idea was that by surveying this range of possible criteria there would be discovered any evident advantage of one over another.

For the kindergarten: In the case of the kindergarten another procedure was necessary. In view of the assumption above, it seemed inconsistent and probably unjustifiable to average the Stanford-Binet and the average Group IQ's in the kindergarten as this would make the Stanford-Binet equivalent to the three group tests used. Dr. Brooks' suggestion was followed here and the kindergarten Criterion III was obtained by averaging Criterion I (Stanford-Binet IQ) taken three times and Criterion II (average Group IQ) taken once. This gives the same relationship as that in the first grade. Criterion IV and Criterion V were then found in the same way as in the other three grades by throwing the criterion first toward Criterion I and then toward Criterion II.

The correlations for these criteria were then found as shown in Table 4, for Criterion I with Criteria II and III, for Criterion II with Criterion III and for Criterion III with Criteria IV and V. An inspection of the results in Table 4 led to the discarding of Criteria III, IV, and V as adding nothing to the other two.

TABLE 4
CORRELATIONS OF THE CRITERIA

Criteria	Kindergarten	First	Second	Third grades
CI, CII	.799	.808	.882	.842
CI, CIII	.977	.943	.959	.944
CIII, CIV	.986	.981	.992	.986
CIII, CV	.982	.985	.984	.995
CII, CIII	.876	.945	.968	.960

The lowest correlation in any of these comparisons of the three criteria is that for correlation of CII, CIII in the kindergarten, which is .876. Aside from this there are only three correlations below .95. These are: In the first grade the correlation of CI with CIII is .943; of CII with CIII is .945; and in the third grade the correlation of CI with CIII is .944. None of these correlations would be regarded as low. All others vary from .959 second grade for CI, CIII to .995 in third grade for CIII, CV. There seemed to be nothing to be gained by using other criteria which correlate so closely with the ones to be used. Therefore in this report the criteria used for correlations are the Stanford-Binet IQ (CI) and the average Group IQ (CII) for the groups of 50 in each grade and for the entire groups in each grade, Criterion I for the groups of 50 and Criterion II for the whole groups.

Another type of criterion was necessary for another purpose. It has been stated that because the Stanford-Binet test was given to only 50 children from each grade, any study made on this basis was limited to about one-fourth of the whole first year group and about one-half of the whole group tested the second year. This would require the assumption that the group of 50 in each grade was a fair sample of the entire group. This was measured for the first year group. Any smaller group may be taken as a fair sample of a larger group if the range of the unit of measurement is the same and if there is a similar spread of units throughout this range. The criteria by which these samplings were measured were the similarities of the Range, Median and Q, of the deviations of the Group IQ's for each test in each grade from Criterion II for the entire group and for the group of 50. These data are presented in Table 5.

It is believed that these data justify the use of the smaller groups as representative of the larger groups for the work of comparison of the tests on the basis of their correlation with the Stanford-Binet IQ.

It seemed desirable to discover what if any strengthening of results would be gained by the use of the whole groups. It is for this reason that Criterion II has been retained. By

TABLE 5

MEDIAN, Q, AND RANGE OF DEVIATIONS OF GROUP IQ'S FROM CII¹
KINDERGARTEN

Tests	175 children			50 children		
	C-V ²	Dn.	P-C	C-V	Dn.	P-C
Med.	5.34	5.82	8.94	4.17	5.00	8.00
Q.	2.82	3.32	5.53	2.03	3.88	5.63
Range	0-18	0-22	0-26	0-17	0-17	0-26

FIRST GRADE

186 Children

Tests	C-V	Dn.	Dt.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Med.	4.79	4.95	5.35	9.80	8.13	6.88	10.36	6.30	7.67
Q.	3.02	2.83	2.95	5.16	4.57	4.18	5.12	3.64	4.23
Range	0-28	0-21	0-28	0-31	0-30	0-58	0-34	0-24	0-25

50 Children

Tests	C-V	Dn.	Dt.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Med.	5.35	5.25	5.50	11.00	10.00	7.33	11.00	7.50	6.80
Q.	2.80	2.98	3.17	5.84	4.04	6.06	4.13	2.98	4.19
Range	1-21	0-16	0-26	0-31	0-29	0-34	1-34	0-22	0-21

SECOND GRADE

235 Children

Tests	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Med.	4.81	11.55	7.35	7.50	8.15	5.07	6.14
Q.	2.42	5.82	3.84	4.25	3.69	2.73	4.04
Range	0-17	0-43	0-31	0-47	0-23	0-19	0-33

50 Children

Tests	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Med.	6.00	13.00	8.75	8.00	8.00	6.00	5.00
Q.	2.10	5.54	3.77	4.63	4.06	2.71	4.59
Range	0-15	1-43	0-24	1-33	0-22	0-17	0-28

THIRD GRADE

204 Children

Tests	Dn.	Hag.	Kg.	Myer	Otis
Med.	5.35	9.36	6.75	7.67	6.38
Q.	2.99	5.09	4.09	4.95	3.25
Range	0-28	0-30	0-30	0-49	0-21

50 Children

Tests	Dn.	Hag.	Kg.	Myer	Otis
Med.	7.00	9.75	7.00	6.83	6.63
Q.	3.83	4.19	4.58	5.90	2.86
Range	0-27	1-23	0-30	0-41	1-17

¹ This table should be read: For the kindergarten with 175 children; for the Cole-Vincent test the median deviation from Criterion II is 5.34, the Q is 2.82 and the range of deviations is from 0 to 18, etc. For purposes of comparison then read: For the kindergarten with 50 children, for the Cole-Vincent test the median deviation from Criterion II is 4.17, the Q is 2.03 and the range of deviations is 0-17. Continue the reading in a similar way for all tests in each group for each grade.

² The test abbreviations should be read as follows:

C-V—Cole-Vincent
Dn.—Dearborn
Dt.—Detroit

Hag.—Haggerty
Kg.—Kingsbury
Myer.—Myers

Otis
P-C—Pintner-Cunningham
Pr.—Pressey

using this criterion both the whole groups and the groups of 50 have been measured. A comparison of the co-efficients of correlation obtained for the two groups has been made with this method of measurement. These co-efficients of correlation are in each case the correlation of tests in the groups of 50 with Criterion I and in the whole groups with Criterion II. In every case the correlation discussed in this study is the partial correlation with chronological age held constant. The comparisons possible will give evidence as to the relative stability of the results of each test when measured in somewhat different ways. A general stability of the results for any test would tend to show the justice of its measurement as found in these groups.

This method gives also a chance to check on the validity of the small group as representative of the large group. If the correlation of the small group which is taken out of the large group should have a reasonable correspondence to the correlation of the large group, the validity of the small group as a fair sample of the large group would be confirmed. If, on the other hand, there was a wide variation in these co-efficients of correlation the results would tend to discredit the use of the small group as representative of the large group. As the data are presented evidence will appear to confirm or discredit the assumption on which this part of the study is based.

CHAPTER 4

THE FIRST YEAR'S STUDY

DATES OF TESTING

The giving of the group intelligence tests to 800 children occupied the time from September 22 to December 12, 1924, a period not quite three months in extent. By checking tests as they came in it was possible at the end of this time to tell what children had taken all the tests for their grade and to make up as many as possible of those lost through one or two absences. From these were taken for each grade 50 children to whom was given the Stanford-Binet Individual Intelligence Test. These children were chosen so as to be for each school, grade, and age, proportional to the number of children of each school, grade, and age tested. The giving of the Stanford-Binet tests continued, except for the Christmas vacation, from December 12, 1924, to the last of February, 1925. The scoring of the tests occupied most of the available time for the remainder of the school year.

THE DATA

Two tests were scored for each child so as to send reports to the school. After that tests were scored only for those who had taken all the tests given to their group. The total number of group tests scored was 5500. The total number of children whose work was available for the first year's study because of their completing all of the tests for their group is as follows:

Grade No. of Children	Kdgtn. 175	First 186	Second 235	Third grades 204	Total 800
--------------------------	---------------	--------------	---------------	---------------------	--------------

The excess in the second and third grades came from the use of whole classes instead of eliminating those not needed for the number it was planned to use. It is mere chance that they balance the deficiency of the lower grades so as to make the total exactly 800. All tests were given and scored by the

writer except in one second grade class where the Pintner-Cunningham test had just been given by the Baltimore Bureau of Research and Statistics. In this instance the scores for this test were kindly furnished by that Department.

A real difficulty arose when the translation of the scores of each test into age equivalents was attempted. Of the nine tests only one, the Cole-Vincent, gave age equivalents for its entire range of scores, and with this test the method of obtaining these equivalents resulted in some questionable norms. In the Cole-Vincent test the age equivalent of a given score is the average of the mental ages measured by the Stanford-Binet test of the group that made that score. This resulted in some parts of the test range in a higher mental age for certain scores than was given to higher scores. This was due, of course, to the tentative nature of the norms at that time. By the time the second year's testing was done this weakness of the test had been corrected and the corrections to correspond were made in the test results for this study. The other tests varied considerably in the degree of completeness with which they gave age equivalents for their entire possible range of scores. This will be shown in Table 6.

TABLE 6
THE RANGE OF TEST SCORES AND THEIR AGE EQUIVALENTS AS
GIVEN BY THE AUTHORS¹

Test	Possible range of scores	Age equivalents provided for scores	Range in ages
Cole-Vincent	0- 50	0- 50	4-6 to 9- 0
Dearborn	0-202	7-179	5-0 to 12-11
Detroit	0- 50	4- 46	4-6 to 7- 6
Haggerty	0-117	30- 80	6-0 to 10- 0
Kingsbury	0-100	22- 56	6-6 to 9- 6
Myers, Grade 1, 2,	0-106	10- 50	6-0 to 14- 0
Grade 3,	0-114		
Otis	0- 85	10- 80	6-0 to 12-11
Pintner-Cunningham	0- 60	8- 49	4-0 to 9- 0
Pressey	0- 80	9- 66	4-6 to 10- 6

¹ This table should be read: For the Cole-Vincent test the possible range of scores is from zero to fifty, age equivalents are provided for the scores zero to fifty, with a possible range of ages in these provided equivalents from four years six months to nine years, etc., with the exception that in the Myers test there is a difference from the other tests in that the first and second grades are supposed not to go so far in one test as the third grade does, and therefore the possible range of scores for them is less. This should be read for the Myers test: the possible range of scores for the first and second grades is from zero to one hundred six and for third grade from zero to one hundred fourteen, age equivalents are provided for the scores ten to fifty, and the range of ages for the equivalents provided is from six years to fourteen years.

It will be seen that while the Detroit, Otis, Pintner-Cunningham and Pressey tests come fairly near to giving age equivalents for their possible range of scores, the Dearborn, Haggerty, Kingsbury and Myers tests give equivalent ages for a relatively short part of their range.

This situation led to trouble in handling the data in this experiment because of the number of instances where a child's score fell beyond, either above or below, the range of scores

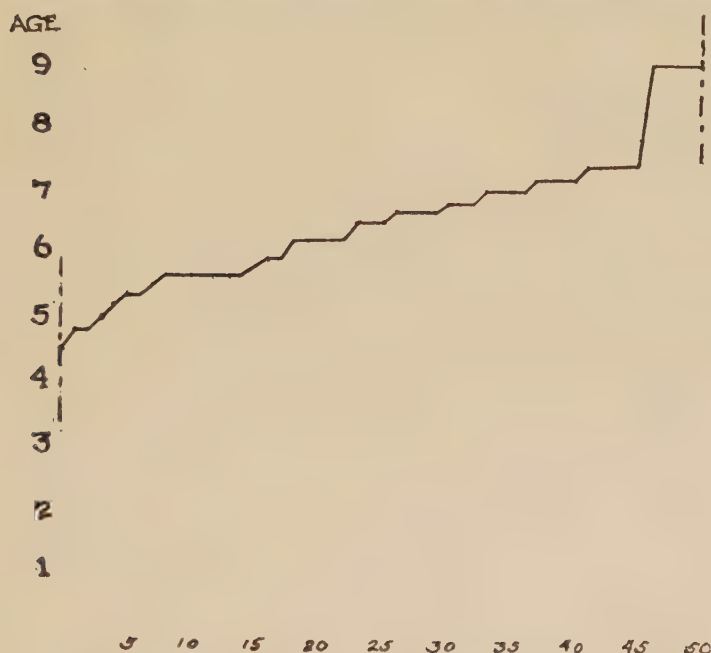


FIGURE I.—Cole-Vincent. Range 0-50. Standards 0-50.
Years 4-6 to 9-0.

for which age equivalents were provided. It was evident that if these tests were to be used some means of adjustment was necessary. Dr. Brooks directed an extension of the scales which would make them usable in these instances. The series of Figures I to IX show the results of plotting the curves of the scores and the age equivalents given, and also the extension of these to cover the possible range of the scores for each test. The short vertical lines cutting the curve in Figures I to IX (pp. 23-29) show in each case the lower and upper

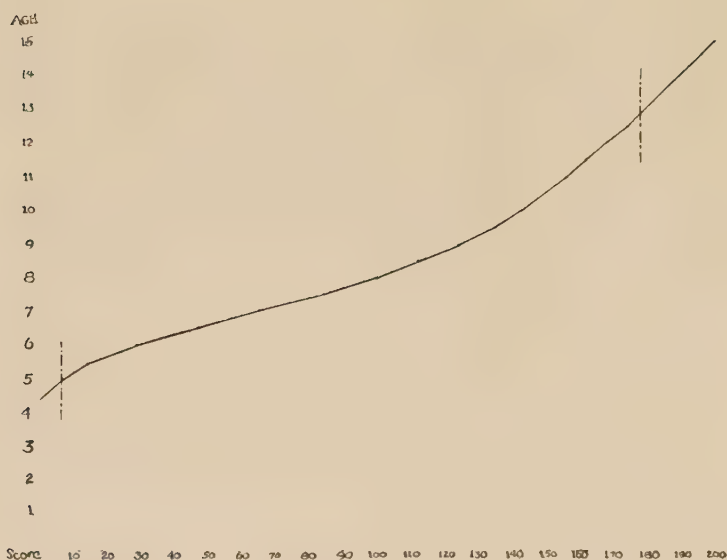


FIGURE II.—Dearborn. Range 0-202. Standards 7-179.
Years 5-0 to 12-11.

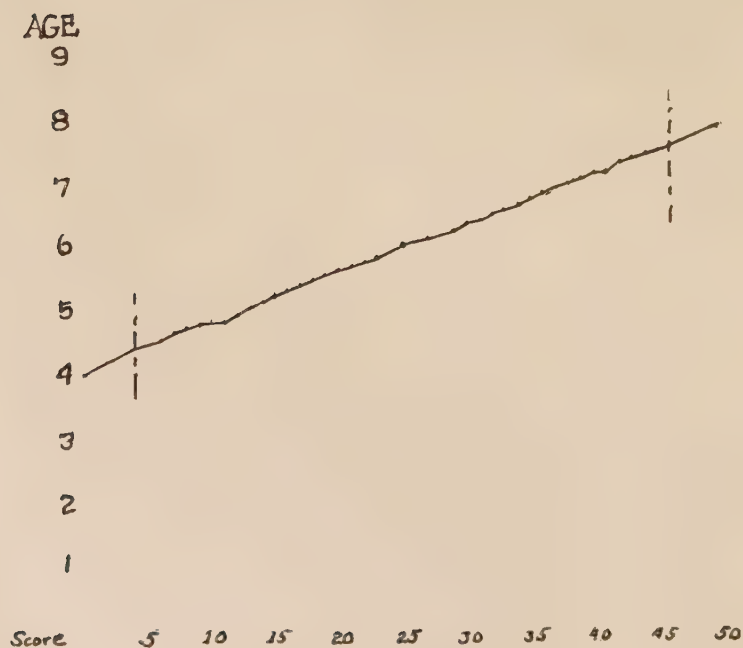


FIGURE III.—Detroit. Range 0-50. Standards 4-46. Years 4-6 to 7-6.

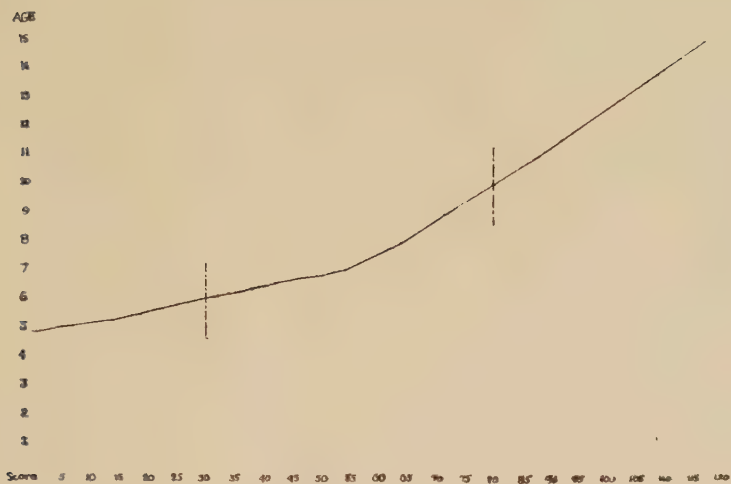


FIGURE IV.—Haggerty. Range 0-117. Standards 30-80.
Years 6-0 to 10-0.

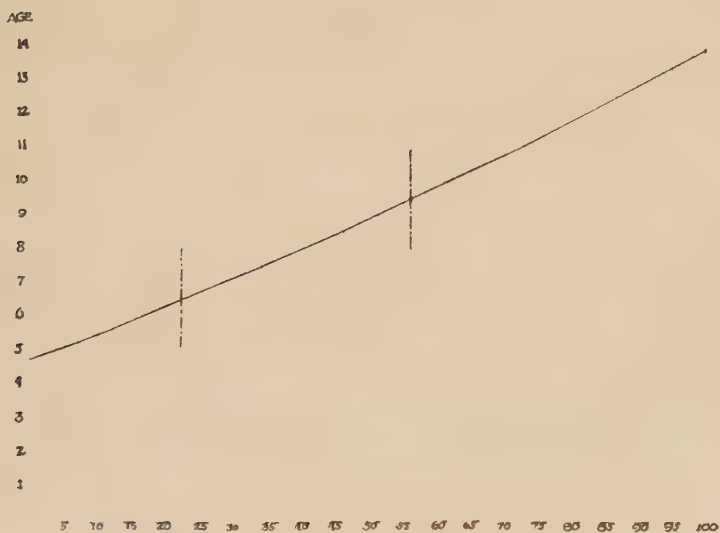


FIGURE V.—Kingsbury. Range 0-100. Standards 22-56.
Years 6-6 to 9-6.



FIGURE VI.—Myers. Range Grade 1, 2-0-106. Grade 3-0-114.
Standards 10-50. Years 6-0 to 14-0.

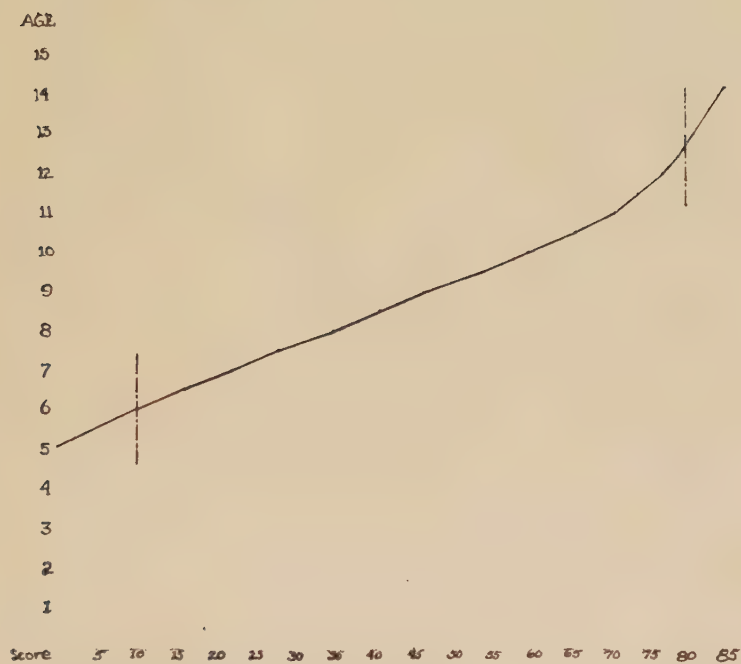


FIGURE VII.—Otis. Range 0-85. Standards 10-80. Years 6-0 to 12-11.

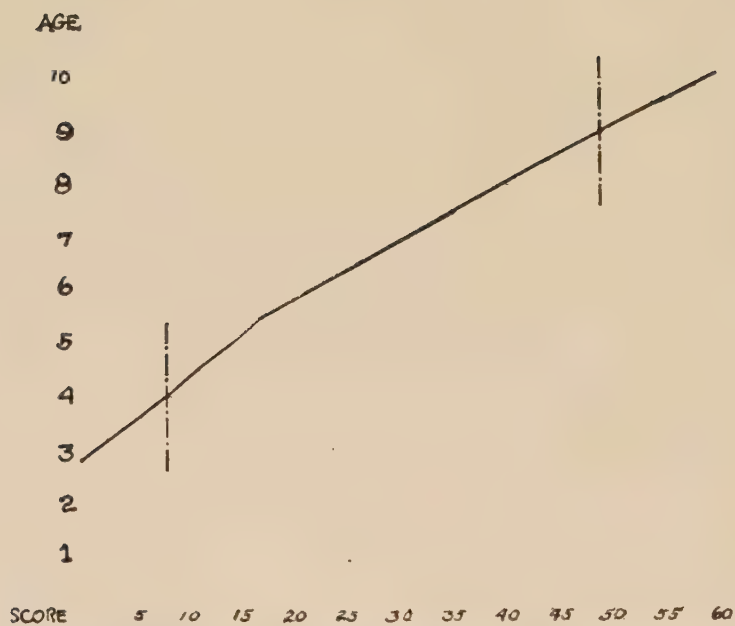


FIGURE VIII.—Pintner-Cunningham. Range 0-60. Standards 8-49.
Years 4-0 to 9-0.

limits of the range covered by the equivalents furnished with the test. A further standardization of the test might show that the curve changes its form beyond the limits given, but in lieu of any evidence to the contrary it seems reasonable to assume that the extension of the curve would be in harmony with the part known. From Figures I to IX, tables of

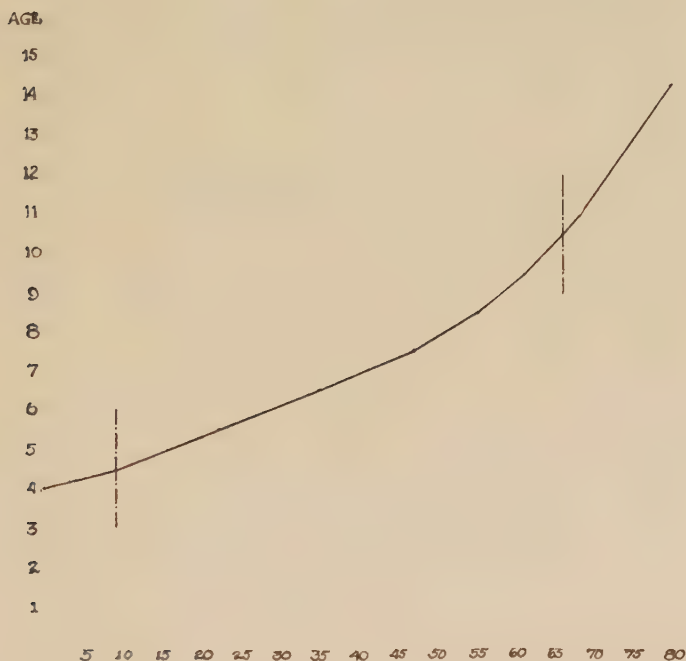


FIGURE IX.—Pressey. Range 0-80. Standards 9-66.
Years 4-6 to 10-16.

equivalent scores were made out, in terms of months of age for given scores, which were as fully as possible in harmony with those furnished with the tests. These are given in Table 7.

In accordance with the age equivalents determined as described, the scores on all tests were translated into mental age terms and from these and the chronological age the Group IQ was found. It is recognized that the Group IQ

TABLE 7—CONCLUDED

		KINGSBURY											
Years	Months	0	1	2	3	4	5	6	7	8	9	10	11
4									0	1	2	3	4
5		5	6	7	8	9	10	11	12	13	14	15	15
6		16	17	18	19	20	21	22	23	24	25	26	27
7		28	29	30	31	32	33	34	35	36	37	38	39
8		40	41	42	43	44	45	46	47	48	49	50	50
9		51	52	53	53	54	55	56	57	58	59	59	60
10		62	63	64		65	66	67		68	69	70	
11		71	72	73		74	75	76	77	78	78	79	80
12		81	82	83	84		85	86	87		88	89	90
13		91	92	93		94	95	96		97	98		99
14		100											

		MYERS											
Years	Months	0	1	2	3	4	5	6	7	8	9	10	11
4									0	1		2	3
5		4		5		6		7		8		9	
6		10	10	11	11	12	12	13	13	14	14	15	15
7		16	16	17	17	18	19	20	20	21	21	22	22
8		23	23	24	24	25	25	26	26	27	27	28	28
9		29	29	30	30	31	31	32	32	33	33	34	34
10		35	35	35	36	36	36	37	37	37	38	38	38
11		39	39	40	40	41	41	42	42	42	43	43	43
12		44	44	44	45	45	45	46	46	46	47	47	47
13		48	48	48	48	48	48	49	49	49	49	49	49
14		50		53		54					55		
15		56		57		58						60	
16		61		62				63		59		64	
17		65			66		67						68
18		69					70					71	
19		72											

		OTIS											
Years	Months	0	1	2	3	4	5	6	7	8	9	10	11
5		0	1	2	3	4	5	6	7	8	9	10	11
6		10	11	12	13	14	15	16	17	18	19	20	21
7		22	23	24	25	26	27	28	30	31	32	33	34
8		35	36	37	38	39	40	41	42	43	44	45	46
9		47	48	49	50	51	52	54	55	56	57	58	59
10		60	61	62	63	64	65	66	67	68	69	70	70
11		71	71	72	72	73	73	74	74	75	75	76	76
12		77	77	78	78	79	79	79	79	79	80	80	80
13		81		82			83					84	
14				85									

		PINTNER-CUNNINGHAM											
Years	Months	0	1	2	3	4	5	6	7	8	9	10	11
2												0	
3		1	2	3			4	5		6		7	
4		8		9		10		11		12		13	
5		14		15		16		17		18	19	20	21
6			22	23		24	25	26		27	28	29	30
7			31	32	33		34	35	36		37	38	39
8			40	41	42		43	44	45		46	47	48
9		49	50	51	52		53	54	55		56	57	58
10		59											

		PRESSEY											
Years	Months	0	1	2	3	4	5	6	7	8	9	10	11
4		3	4	5	6	7	8	9	10	11	12	13	14
5		15	16	17	18	19	20	22	23	24	25	26	27
6		29	30	31	32	33	34	35	36	37	38	39	40
7		41	42	43	44	45	46	47	48	48	49	50	50
8		51	52	52	53	54	54	55	56	56	57	57	58
9		58	59	59	60	60	61	61	61	62	62	63	63
10		64	64	64	65	65	65	66	66	66	67	67	67
11		68			69			70				71	
12			72				73			74			75
13				76				77				78	
14		79	79				80						

KINDERGARTEN, FIRST YEAR

A group of 175 children in the kindergarten took the three group tests given in this grade. The statistical data from the various calculations with the results of these tests have been accumulated in Tables 8 to 10.¹ In these tables the results for the three tests are arranged in such a way as to show rank order of the tests beginning with the one which stands the highest. Table 8 is derived from Table 5, which was discussed in Chapter 3. It is introduced here to make convenient a comparison of the ranks of the tests by the different methods of measurement. Table 9 gives the correlations for each test with the two criteria, their PE values and $1-k$ which was obtained from Miner's tables.²

TABLE 8
RANK BY MEDIAN DEVIATION OF IQ FROM CII¹
50 CHILDREN 175 CHILDREN

Test	Rank	Med. dev.	Q	Test	Rank	Med. dev.	Q
Cole-Vincent	1	4.17	2.03	Cole-Vincent	1	5.34	2.82
Dearborn	2	5.00	3.88	Dearborn	2	5.82	3.32
Pintner-Cunningham	3	8.00	5.63	Pintner-Cunningham	3	8.94	5.53

¹ This table should be read: The Cole-Vincent test ranks first with a median deviation of 4.17 and a Q of 2.03, when ranked by median deviation and Q from CII, etc.

In Table 8 it will be seen that the rank of the tests as measured by the median deviation of the GIQ's from Criterion II, is the same for the group of 50 as for the group of 175 and that by this measure the Cole-Vincent test stands first, the Dearborn test, second, and the Pintner-Cunningham test, third. The median deviation of the Pintner-Cunningham test is nearly double that of the Cole-Vincent in the group of 50. Because there is less conformity between the large and small groups for the Cole-Vincent test, there is not

¹ The basic tables and work sheets upon which the data presented in this study depend are on file in the library of The Johns Hopkins University.

² Miner, J. R., Tables of $\sqrt{1-r^2}$ and $1-r^2$ for Use in Partial Correlation and in Trigonometry. The Johns Hopkins Press, Baltimore, Maryland. 1922.

so great a discrepancy between the Cole-Vincent and the Pintner-Cunningham tests in the group of 175. However, it is significant that the rank is the same.

TABLE 9
RANK OF TESTS BY PARTIAL CORRELATION¹
50 CHILDREN CI

Test	Rank	r	PE	1-k
Pintner-Cunningham	1	.803	.034	.404
Cole-Vincent	2	.756	.041	.345
Dearborn	3	.643	.056	.234

175 CHILDREN CII

Test	Rank	r	PE	1-k
Pintner-Cunningham	1	.944	.006	.670
Cole-Vincent	2	.888	.011	.540
Dearborn	3	.799	.018	.399

¹ This table should be read: The Pintner-Cunningham test ranks first with r .803, PE .034, and $1-k$.404, etc. The measure $1-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .404.

In Table 9 there are given for comparison the partial correlations with PE and $1-k$ for the 50 children with Criterion I and the partial correlations with PE and $1-k$ for the 175 children with Criterion II. Here it will be seen that as measured by the method of correlation the tests hold the same rank in the two groups with different criteria. It will be well to recall that Criterion I is not directly related to the group test data while Criterion II is the average of the Group IQ's of the three tests. Also it will be worth while to note that the correlation between Criterion I and Criterion II is .799. If then the correlation of the small group with the external criterion and of the large group which contains the small group with the criterion derived from the test data, gives the same relative position to the tests there is probably some significance in the result. However, a situation which is, at least on its face, disconcerting is present. While it was found that measured by the median deviation from Criterion II the rank of the tests in order of merit was Cole-Vincent, Dearborn, Pintner-Cunningham, here where the value is measured by the correlation with Criterion I and Criterion II the rank is Pintner-Cunningham, Cole-Vincent, Dearborn.

That is, while Cole-Vincent and Dearborn do not change their relative positions, Pintner-Cunningham moves from last place to first. The difference in correlation is not very great, .05 in the group of 50 and .056 in the group of 175, but it is in a part of the range where considerable significance attaches to the difference. That is, a correlation of .80 (in the group of 50) is regarded with more respect than a correlation of .75. Also the co-efficient $r-k$ shows that the Pintner-Cunningham test, r .803, gives a 40% better guess than chance, while the Cole-Vincent test with r .753 gives only 34% better than chance guess in estimating intelligence.

The question of attempting to harmonize the two rankings or to explain the discrepancy needs attention. There is a possible explanation based on the meaning of the two methods of measurement. The method of correlation shows how nearly the tests used have ranked the subjects in the same order as is done by a given criterion. For this situation that means that the Pintner-Cunningham test comes nearer than the other two tests to placing the children in the same rank as to intelligence as the Stanford-Binet test does, and also that it agrees more closely in its ranking with the rank which would obtain if all three tests were given and the average IQ taken as the basis of the ranking. The method of median deviation shows as the term implies what will be the median of the deviations from the criterion of the standing of the subjects tested. A smaller median deviation means a closer agreement of the test with the criterion as indicated by this measure. By reference to Table 5, it will be seen that the rank by Q is the same as the rank by the median deviation. This means that there is a smaller spread of the middle 50% of cases about the smaller median deviation for the Cole-Vincent test than for the other two. It is probable that individual children will be measured more exactly by the test having the smallest median deviation and Q . Therefore, the possible significance of the differences between the two methods of measuring the tests lies in the purpose one has in testing. If one desires merely to rank the children so as to place them in similar ability groups then it would

seem to be wise to use the test which correlates most highly with the criterion, in this case the Pintner-Cunningham test. If what one wants is to get the measure of individual children which agrees most closely with their measure by the criterion then it would seem best to use the test which has the smallest median deviation and Q from the criterion, in this case the Cole-Vincent test.

TABLE 10
BEST CHOICE OF TESTS—50 CHILDREN
SINGLE—RANK BY CI¹

Test	Rank	r	PE	1-k
Pintner-Cunningham	1	.803	.034	.404
Cole-Vincent	2	.753	.041	.342
Dearborn	3	.643	.056	.234

COMBINED IN PAIRS—RANK BY CI¹

Test	Rank	r	PE	1-k
Dearborn and Pintner-Cunningham	1	.873	.023	.512
Cole-Vincent and Pintner-Cunningham	2	.826	.030	.436
Cole-Vincent and Dearborn	3	.706	.049	.292

¹ This part of table is a duplicate of the first part of Table 9 and should be read in the same way.

² This table should be read: The mean of the Dearborn and Pintner-Cunningham tests ranks first with r .873, PE .023 and $1-k$.512, etc. The measure $1-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .512.

Table 10 shows the partial correlation, PE and $1-k$ of the Mean IQ's of the tests combined in pairs, with Criterion I. The part of the Table 9 giving the partial correlation of the single tests with Criterion I is copied to make comparison easier. When these tests are combined by giving equal weight to each it is evident that the use of the Pintner-Cunningham test and Dearborn test together excels the use of either alone. The combination gives r .873, while the Pintner-Cunningham test alone has r .803 and the Dearborn test is r .643. When $1-k$ is considered it is seen that we improve the prophecy over chance by the combination 51%, by Pintner-Cunningham alone, 40%, and by Dearborn alone, 23%. That is, the combination is 11% better than Pinter-Cunningham alone and more than twice as good as Dearborn alone.

The combination of the Cole-Vincent and Pintner-Cunningham tests in the same way gives very slight gain over Pintner-Cunningham alone (.023), while the difference in $1-k$ is .44—.40. This indicates that one would gain only 4% in accuracy by adding the Cole-Vincent test to the Pintner-Cunningham test. When Cole-Vincent is considered the combination gives an increase in correlation of .073, and when $1-k$ is considered the gain is .44—.34 or 10%. This indicates that for groups comparable to those studied it would add 10% to the accuracy of measurement of intelligence to add the Pintner-Cunningham test to the Cole-Vincent.

The other possible combination of tests is the Cole-Vincent and Dearborn tests. This combination is not so satisfactory as either of the others. It gives r .706. This is lower than the Cole-Vincent alone which indicates at once the undesirability of making such a combination. The combination improves the estimate made by chance only 29% while the Cole-Vincent alone improves a chance estimate of intelligence 34%. The combination is better than the Dearborn alone by .063 and the difference in $1-k$ value is .29—.23 or 6%. The apparent conclusion here is that one will gain little by using the Dearborn test in the kindergarten for purposes of ranking in harmony with the Stanford-Binet unless it is supplemented by the Pintner-Cunningham test. The question then to decide is whether a gain of 11% in the reliability of the rating, as measured by the difference in $1-k$ value is sufficient to warrant the additional time, effort and expense of using the Dearborn test.

Summary

1. The Cole-Vincent test ranks first, the Dearborn test ranks second, and the Pintner-Cunningham test ranks third when measured by the median deviation and Q from Criterion II. This shows relative degree of conformity of individual measures to the criterion.

2. The Pintner-Cunningham test ranks first, Cole-Vincent, second, and Dearborn, third, when measured by correlation with either Criterion I for the group of 50 or with Criterion

II for the group of 175. This shows relative value in obtaining similar rank order, by the use of each test, to those obtained by the criterion.

3. Combining these tests in various ways, it is found that Dearborn and Pintner-Cunningham rank first, Cole-Vincent and Pintner-Cunningham, second, and Cole-Vincent and Dearborn, third. This last combination is contra-indicated as being disadvantageous to one of the pair and giving little additional value to the other.

FIRST GRADE, FIRST YEAR

The results of the various calculations with the first grade data for 186 children are accumulated in Tables 11 to 13.¹ In these tables the data are arranged so as to show the rank order of the tests when measured by different criteria.

TABLE 11

RANK BY MEDIAN DEVIATION OF IQ FROM CII¹
50 CHILDREN 186 CHILDREN

Test	Rank	Med. dev.	Q	Test	Rank	Med. dev.	Q
Dearborn	1	5.25	2.98	Cole-Vincent	1	4.78	3.02
Cole-Vincent	2	5.33	2.80	Dearborn	2	4.95	2.83
Detroit	3	5.50	3.17	Detroit	3	5.35	2.95
Pressey	4	6.80	4.19	Pintner-Cunningham	4	6.30	3.64
Myers	5	7.33	6.06	Myers	5	6.88	4.18
Pintner-Cunningham	6	7.50	2.98	Pressey	6	7.67	4.23
Kingsbury	7	10.00	4.04	Kingsbury	7	8.13	4.57
Otis	8½	11.00	5.84	Haggerty	8	9.80	5.16
Haggerty	8½	11.00	4.13	Otis	9	10.36	5.12

¹ This table should be read: The Dearborn test is first in rank with the median deviation 5.25 and Q 2.98, etc.

Table 11 shows the rank of the tests as measured by the median deviation of the GIQ's of the tests from Criterion II for the group of 50 and for the group of 186 which includes the group of 50. This is taken from Table 5 in Chapter 3. For the group of 186 children this rank order is Cole-Vincent, Dearborn, Detroit, Pintner-Cunningham, Myers, Pressey, Kingsbury, Haggerty, Otis, with a range of median deviation

¹ For statement regarding original data see page 32.

from 4.78 to 10.36. The order for the 50 children is Dearborn, Cole-Vincent, Detroit, Pressey, Myers, Pintner-Cunningham, Kingsbury, Otis, Haggerty, with a range of median deviation from 5.25 to 11.00. The order is changed for the selected group from that of the whole group by the exchange in rank of Cole-Vincent and Dearborn and of Pintner-Cunningham and Pressey. In the first case one step displacement is made in each test. For the group of 50 the difference in median deviation is .08 in favor of the Dearborn test. For the group of 186 the difference is .16 in favor of the Cole-Vincent test. This is probably not a serious amount of disagreement between the ranking in the two groups. In the case of the Pintner-Cunningham and Pressey tests the change in rank of each is two places. The difference in the median deviation of the two tests for the group of 50 is .70 in favor of the Pressey test. In the group of 186 the difference is 1.37 in favor of the Pintner-Cunningham test. This is an amount of difference which is not so insignificant as that in the other two tests. A search for the cause of this difference shows that the median deviation for the Pressey test is 7.67 for the group of 186 and 6.80 for the group of 50, giving a difference of .87 in favor of the group of 50, while the median deviation for the Pintner-Cunningham test is 6.30 for the group of 186 and 7.50 for the group of 50, giving a difference of 1.20 in favor of the group of 186. That is, the selected group of 50 appears to be a less satisfactory representative of the larger group when both are measured by the Pintner-Cunningham and the Pressey tests than when measured by others of the tests. It would be interesting to attempt to find the cause for this. Every other test, five in number, holds the same rank order in the two groups. It is also true that when the range of these ranks is divided into thirds there is no difference in the two groups in the placement of any test, that is, each test is in the same third of the range of the ranks when measured in either group. Because of the closeness of similarity in this ranking, the ranking by the group of 50 will tentatively be accepted because that is the group with which the intensive study is made. The differences between

these two groups will be held in mind as possibly needed to check either by corroboration or by refutation other measurements found.

TABLE 12
RANKS OF TESTS BY PARTIAL CORRELATION¹
50 CHILDREN CI

Test	Rank	r	PE	r-k
Detroit	1	.792	.036	.390
Pintner-Cunningham	2	.720	.046	.306
Otis	3	.663	.053	.251
Cole-Vincent	4	.655	.054	.244
Myers	5	.654	.055	.244
Dearborn	6	.650	.055	.240
Pressey	7	.626	.058	.220
Haggerty	8	.591	.062	.193
Kingsbury	9	.545	.067	.162

186 CHILDREN CII

Test	Rank	r	PE	r-k
Otis	1	.915	.008	.597
Pintner-Cunningham	2	.887	.011	.538
Pressey	3	.885	.011	.534
Kingsbury	4	.841	.015	.459
Myers	5	.835	.015	.450
Dearborn	6	.825	.016	.435
Detroit	7	.822	.016	.431
Cole-Vincent	8	.802	.018	.403
Haggerty	9	.749	.022	.337

¹ This table should be read: The Detroit test ranks first with r .792, PE .036, $r-k$.390, etc. The measure $r-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .390.

Table 12 gives the ranks of the tests by partial correlation (chronological age constant) for the group of 50 with Criterion I and for the group of 186 with Criterion II. It also gives the PE value and $r-k$ for these correlations. The ranking for these two groups is as follows:

50 CHILDREN

1. Detroit
2. Pintner-Cunningham
3. Otis
4. Cole-Vincent
5. Myers
6. Dearborn
7. Pressey
8. Haggerty
9. Kingsbury

186 CHILDREN

1. Otis
2. Pintner-Cunningham
3. Pressey
4. Kingsbury
5. Myers
6. Dearborn
7. Detroit
8. Cole-Vincent
9. Haggerty

It should be noted that the correlations in Table 12 are considerably higher with Criterion II which is the average of the GIQ's than with Criterion I which is the IQ of the Stanford-Binet test. For this reason the IQ's of the two groups cannot themselves be compared. The relative positions of the tests in the groups can however be compared. This is shown above. Taking the group of 50 as the basis of comparison this shows the following displacement in rank:

Test	Change in rank	Displacement
Detroit	1-7	-6
Pintner-Cunningham	2-2	0
Otis	3-1	+2
Cole-Vincent	4-8	-4
Myers	5-5	0
Dearborn	6-6	0
Pressey	7-3	+4
Haggerty	8-9	-1
Kingsbury	9-4	+5

This shows, for example, that while the Detroit test gives the highest correlation with Stanford-Binet test it is relatively low (Rank 7) in its correlation with the average of all the tests. A comparison of its 1-k value in the two rankings is interesting in view of the wide differences between the two. In the group of 50 the accuracy of judgment over chance, as compared with the Stanford-Binet, is increased 39% by the use of this test; in the group of 186 the improvement over chance judgment when compared with the average of nine group tests is 43%. That is, while there is a wide range in rank the difference in advantage in the use of the test as shown by the two groups with different criteria is less (only 4%) than would be expected from the ranks. The Pintner-Cunningham, Dearborn and Myers tests hold the same rank in both groups. The Haggerty test is displaced only one step, from rank eight to rank nine. Otis is displaced two steps. Cole-Vincent and Pressey are each displaced four places and Kingsbury is displaced five steps. The correlation between Criterion I and Criterion II on the bases of which these ranks are found is for first grade .808. None of the tests correlate as high as this with Criterion I but the Detroit test approximates it with a correlation of .792. Only two tests fall below

the correlation of Criterion I and Criterion II in their correlation with Criterion II. These are the Cole-Vincent, slightly, $r .802$, and the Haggerty, $r .749$. The question as to what significance should attach to these differences in rank is a matter for refined statistical treatment.

In considering Tables 11 and 12 it should be remembered as pointed out in the discussion of the work of the kindergarten¹ that the two methods of handling the data are fundamentally different in significance so that the difference in ranking by the two methods has significance rather when considered in the light of purpose for which the test is to be used than when regarded abstractly.

With so large a number of tests as is used in this grade location in the upper half of the ranks as contrasted with location in the lower half has some interest. On this basis it is found that the upper half (five tests) includes for the group of 50 measured by the median deviation, Cole-Vincent, Dearborn, Detroit, Pintner-Cunningham and Myers; for the 186 children measured by the median deviation Dearborn, Cole-Vincent, Detroit, Pressey and Myers, there being only one change. It appears that so far as these tests are concerned, if one wishes to get exactness in rating individual children in comparison with their rating as it would be on the basis of Criterion II these five tests would be relatively desirable. Possibly it would be safer to give preference to the first three.

On the basis of correlation with Criterion I the tests in the first five ranks are Detroit, Pintner-Cunningham, Otis, Cole-Vincent and Myers. On this basis three of these agree with the rating by median deviation: the Detroit, Cole-Vincent and Myers. On the basis of Criterion II these tests are the first five, Otis, Pintner-Cunningham, Pressey, Kingsbury and Myers. The Otis, Pintner-Cunningham and Myers are the tests which with both criteria rank in the upper half. Other things being equal the external criterion would be accepted as a more valid measure of the worth of a test than one in which the test itself forms a part, in this case, one-ninth. An

¹ Page 34.

interesting comparative study would be a change in Criterion II, making it the median GIQ instead of the average. This would take less account of the extreme cases and would probably show rather definitely any general tendency to a regularity of relationship between the tests.

The tests which show the most correspondence as measured by Criteria I and II are the Pintner-Cunningham, having Rank 2 in both, the Myers and Dearborn which maintain respectively Ranks 5 and 6 with both criteria, the Otis, having Rank 3 with Criterion I and Rank 1 with Criterion II, the Haggerty, having Rank 8 with Criterion I and Rank 9 with Criterion II. The Haggerty also ranks lowest or next to lowest when measured by median deviation. This gives two tests consistently in the upper third of the ranks, two consistently in the middle third of the ranks, and two consistently in the lower third.

Table 13 shows the partial correlations with Criterion I (chronological age constant) of the means of the GIQ's of all possible pairs of tests, and for the 13 pairs ranking highest and the five pairs ranking lowest shows the PE value of the co-efficient of correlation and the 1-k for each co-efficient.

Table 14 shows the same correlations in another arrangement to make easy the comparison of combinations of each test with all others. In this table the rank order is, of course, not shown. One fact that stands out from a comparison of this table with the part of Table 12 which shows correlations with Criterion I, is that the use of the Pressey test is an advantage to every test with which it is combined and is itself also improved by combination with every other test. This is the only one of these tests of which this is true. The Cole-Vincent test improves by combination the Dearborn, Haggerty, Kingsbury, Myers, Pintner-Cunningham, Pressey, but not the Detroit and Otis. It is itself improved by the Detroit, Kingsbury, Myers, Pintner-Cunningham, Pressey, but not by the Dearborn, Haggerty or Otis. Clearly the Otis and Cole-Vincent tests should not be combined since neither helps the other. It would seem better to combine only such tests as are each improved by the combination, in this case

Kingsbury, Myers, Pintner-Cunningham or Pressey with the Cole-Vincent. The Dearborn test improves by combination

TABLE 13

FIRST GRADE—FIRST YEAR—50 CHILDREN. CORRELATION (CA CONSTANT) OF MEAN IQ'S ON PAIRS OF GROUP INTELLIGENCE TESTS WITH CRITERION I

Tests	Rank	r	PE	i-k
Detroit and Pressey ¹	1	.878	.022	.521
Haggerty and Pressey	2	.844	.027	.464
Myers and Pressey	3	.835	.029	.450
Detroit and Otis	4	.827	.030	.438
Detroit and Pintner-Cunningham	5 $\frac{1}{2}$	.821	.031	.429
Pintner-Cunningham and Pressey	5 $\frac{1}{2}$	.821	.031	.429
Dearborn and Pressey	7	.820	.031	.428
Kingsbury and Pressey	8	.816	.032	.422
Cole-Vincent and Pressey	9	.815	.032	.420
Kingsbury and Pintner-Cunningham	10	.804	.034	.406
Otis and Pintner-Cunningham	11	.801	.034	.406
Detroit and Myers	12	.800	.034	.400
Otis and Pressey	13	.799	.035	.399
Myers and Pintner-Cunningham	14	.790		
Dearborn and Myers	15	.777		
Cole-Vincent and Myers	16	.774		
Myers and Otis	17	.768		
Kingsbury and Otis	18	.767		
Dearborn and Otis	19	.766		
Dearborn and Detroit	20	.761		
Cole-Vincent and Pintner-Cunningham	21	.755		
Haggerty and Myers	22	.747		
Haggerty and Otis	23	.743		
Cole-Vincent and Detroit	24	.739		
Cole-Vincent and Kingsbury	25	.736		
Haggerty and Pintner-Cunningham	26	.730		
Kingsbury and Myers	27	.729		
Dearborn and Pintner-Cunningham	28	.724		
Dearborn and Kingsbury	29	.714		
Detroit and Kingsbury	30	.708		
Detroit and Haggerty	31	.690		
Haggerty and Kingsbury	32	.687	.053	.273
Cole-Vincent and Dearborn	33	.653	.055	.243
Cole-Vincent and Haggerty	34	.645	.056	.236
Dearborn and Haggerty	35	.644	.056	.235
Cole-Vincent and Otis	36	.622	.058	.217

¹ This table should be read: The combination of the Detroit and Pressey tests ranks first with r .878, PE .022, and i-k .521, etc. The measure i-k indicates the improvement in estimate by a test over an estimate by chance, which in this test is .521.

the Haggerty, Kingsbury, Otis, Pintner-Cunningham and Pressey, but not the Cole-Vincent, Detroit or Myers. It is improved by the Cole-Vincent, Detroit, Kingsbury, Otis,

Pintner-Cunningham, Pressey. The combinations suggested as desirable for the Dearborn are those with the Kingsbury, Otis, Pintner-Cunningham, Myers and Pressey. In the same way the relationships of each test to its combinations can be picked out from a comparison of Table 14 with Table 12. The Haggerty, Kingsbury, Myers, Pintner-Cunningham, and Pressey tests are improved by each of the other tests, and the Otis by all but the Cole-Vincent.

TABLE 14
FIRST GRADE—FIRST YEAR—50 CHILDREN. CORRELATION (CHRONOLOGICAL AGE CONSTANT) OF THE MEAN IQ'S ON PAIRS OF GROUP INTELLIGENCE TESTS WITH CRITERION I

Tests	Dn.	Dt.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Cole-Vincent	.653	.739	.645	.736	.774	.622	.755	.815
Dearborn		.761	.644	.714	.777	.766	.724	.820
Detroit			.690	.708	.800	.827	.821	.878
Haggerty				.687	.747	.743	.730	.844
Kingsbury					.729	.767	.804	.816
Myers						.768	.790	.835
Otis							.801	.799
Pintner-Cunningham								.821

Table 13 shows these results in rank order so that it is possible to determine readily what tests, on the basis of this study, give the best promise of satisfactory rating of intelligence among first grade children. Since there are 36 possible ranks for these combinations one may first consider the highest of all and then compare those in the upper half of the ranks with those in the lower half of the ranks. The Detroit and Pressey combined give the high correlation of .878, with a 1-k of .52. The correlations then decrease slowly through the various combinations of tests until Rank 13 is reached. The reason for making a division at this point is that the ranks 1-13 include all the combinations of the Pressey test with any other test and as this test seemed to be of value in combination this division was noted to emphasize this value. The lowest correlation in this group, that of Otis and Pressey is .799, or practically .80. With this part of the range while the correlation drops from .878 to .799, the advantage over chance ranking, as indicated by 1-k drops from 52% to 40%

which suggests a wide range in the desirability of these combinations. Looking for a moment at the lower end of this range of ranks it will be seen that for the lowest five pairs the range of correlations is from .687 to .622 with a range of 1-k from 27% to 22%, showing a much smaller advantage to be gained by using these test combinations than by using those in the upper range of the ranks.

Another feature that appears in a comparison of parts of this table is the various locations of certain tests in different combinations. Take for example, the Detroit test which ranks first in combination with the Pressey. In combination with the Dearborn it is thrown into the lower half of the range with a rank of 20; in combination with Cole-Vincent it ranks 24; in combination with Kingsbury it ranks 30; with Haggerty, it ranks 31. There are very clear suggestions here as to contra-indicated combinations. The Pressey, which ranks first with the Detroit, has its lowest rank 13. It remains well in the upper half of the range and may be regarded as a satisfactory combiner. Other tests may be studied in the same way. An interesting comparison is that which shows the number of times each test appears in the upper and in the lower half of this range of ranks. This is shown in Table 15.

TABLE 15

NUMBER OF TIMES EACH TEST APPEARS IN COMBINATION IN THE
UPPER HALF OF THE CORRELATIONS AND IN THE LOWER
HALF OF THE CORRELATIONS

Half	C-V	Dn.	Dt.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Upper	2	2	4	1	3	6	5	5	8
Lower	6	6	4	7	5	2	3	3	0

This suggests that next to the Pressey test the Myers test makes the best combination with other tests in the first grade. Its satisfactory combinations are with the Pressey, Detroit, Pintner-Cunningham, Cole-Vincent, Otis, and Dearborn tests. It is interesting that while in the upper half of these correlations there is no correlation as low as .75, in the lower half there are only three which are above this point.

Many persons interested in the statistical study of measurements regard .60 as an acceptable correlation. On this basis

any of these combinations would be satisfactory for use. But it is important to note the relative position above this theoretically acceptable point. From the standpoint of prediction this point is not so good as may be seen from the r - k values of the lower end of this range. The mean co-efficient of correlation is .759. This shows that if this range of co-efficients of correlation is divided into halves according to the degree of correlation there are 20 pairs in the upper half and 16 in the lower half of the range.

This range of ranks assigned to given tests when combined with different tests is interesting. This is shown in Table 16.

TABLE 16
RANGE IN RANKS OF TESTS WHEN COMBINED WITH OTHER TESTS

Test	Range
Pressey	1 -13
Detroit	1 -31
Haggerty	2 -35
Myers	3 -27
Otis	4 -36
Pintner-Cunningham	5½ -28
Dearborn	7 -35
Kingsbury	8 -32
Cole-Vincent	9 -36

The lowest end of the range in each case is the first appearance of the test in combination. This gives some idea of the ranking of the test. It will be seen that the Pressey test is the only test which is confined in its combinations to the upper half of the range whether considered as the range of ranks or as the range of correlation co-efficients.

Summary

With three times the number of tests in the first grade that were used in the kindergarten the comparison is much more complicated and provides for many more opportunities for discrepancies which call for explanation. A thorough study of these would require much more time than is available for such work in this report. An effort will be made to state the situations as they appear with regard to these tests,

and to call attention to any indications of conclusions which may be drawn:

1. Cole-Vincent, Dearborn and Detroit tests are given the first three ranks by the median deviation with either group of children. Pintner-Cunningham, Myers and Pressey are by the same criterion placed in the middle three ranks. Kingsbury, Otis and Haggerty are placed in the three lowest ranks by this criterion in both the group of 50 and the group of 186. This on the whole gives agreement in the ranking.

2. There is agreement in these rankings when the ranking is by the correlation of the GIQ's with Criterion I¹ in the following cases:

Cole-Vincent and Detroit are placed in the upper half as they are by the median deviation.

Kingsbury and Haggerty are placed in the lower half of the ranks by Criterion I as they are by the median deviation.

3. The correlation of the GIQ's with Criterion II for the 186 children, when the range is divided into halves, agrees with the ranking by the median deviation for the 186 children in the following cases:

Pintner-Cunningham and Myers are placed in the upper half by both criteria.

Haggerty is placed in the lower half by both criteria.

When compared with the group of 50 ranked by the median deviation the agreements are as follows:

Pressey and Myers tests are placed in the upper half by both rankings.

Haggerty is placed in the lower half by both rankings.

4. There is agreement in placing Pintner-Cunningham and Otis tests in the upper half of the range and in placing Haggerty in the lower half of the range when the ranking by partial correlation is compared for the group of 50 with Criterion I and the group of 186 with Criterion II.

5. When these two groups are divided into thirds for comparison it is found that:

¹ Attention is called to the fact that throughout this study the correlations discussed are partial correlations with chronological age held constant.

Otis and Pintner-Cunningham are common to the upper third in the two groups.

Myers and Dearborn are common to the two groups in the middle third.

Haggerty is in the lowest third in the two groups.

This agrees in only two instances with the comparison by thirds of the ranking by median deviation. Myers test is in the middle group as ranked by both methods.

Haggerty test is in the lowest third as ranked by both.

6. An effort to find advantageous combinations of tests by pairing the tests studied shows a variation in value as indicated by the correlation with Criterion I of the mean GIQ's of all possible pairs of tests, from $r .878 \pm .022$ to $r .608 \pm .060$. (Table 13.)

Pressey test was found to be of value in combination with every other test.

The Cole-Vincent and Dearborn tests only appear once in the first half of the ranks besides their combination with the Pressey test.

The Haggerty test does not appear in the first eighteen of the 36 possible ranks except in combination with Pressey. In this combination it is interestingly high, ranking second.

The tests, which by their more frequent location in the upper half of the range of ranks, give the most promise are:

Pressey appears only in the first 13 ranks.

Myers appears six times in the first 18 ranks.

Pintner-Cunningham and Otis tests each appear five times in the first 18 ranks.

Detroit appears four times in each half.

SECOND GRADE, FIRST YEAR

The results of the various calculations with the second grade data for 235 children are accumulated in Tables 17 to 20.¹ In these tables the data are arranged to show the rank order of the tests when measured by different criteria.

¹ For statement regarding original data see page 32.

Table 17 shows the rank of the tests as measured by the median deviation of the GIQ's from Criterion II for the group of 50 and for the group of 235 which includes the group of 50. This is taken from Table 5, Chapter 3. For the group of 50 children this rank order is Pressey, Dearborn and Pintner-Cunningham holding the same rank, Otis and Myers holding the same rank, Kingsbury, Haggerty. The range of median deviations is from 5.00 to 13.00. The order for the 235 children is Dearborn, Pintner-Cunningham, Pressey, Kingsbury, Myers, Otis, Haggerty, with a range of median deviations from 4.81 to 11.55. The change in order of rank for the selected group is as follows: Dearborn falls

TABLE 17
RANK BY MEDIAN DEVIATION OF IQ FROM CRITERION II
50 CHILDREN 235 CHILDREN

Test	Rank	Med. dev.	Test	Rank	Med. dev.
Pressey	1	5.00	Dearborn	1	4.81
Dearborn	2½	6.00	Pintner-Cunningham	2	5.07
Pintner-Cunningham	2½	6.00	Pressey	3	6.14
Otis	4½	8.00	Kingsbury	4	7.35
Myers	4½	8.00	Myers	5	7.50
Kingsbury	6	8.75	Otis	6	8.15
Haggerty	7	13.00	Haggerty	7	11.55

¹ This table should be read: The median deviation of the GIQ's of the Pressey test from Criterion II is 5.00, etc.

from first to second place; Pressey rises from third to first place; Kingsbury falls from fourth to sixth place; Myers rises from fifth place to Rank 4½. Pintner-Cunningham and Haggerty hold the same position in both groups, the Pintner-Cunningham test being in second rank and the Haggerty test being in seventh rank in both. Since in the ranking for the group of 50 the tests fall into five different ranks a convenient method of comparison is to divide the range of ranks into the higher three and the lower four. On this basis the same three tests, Pressey, Dearborn and Pintner-Cunningham, have the first three ranks in the two groups. In the lower four Kingsbury and Otis exchange places, Myers and Haggerty maintain the same positions. There seems to be a rather satisfactory consistency in the measurement of these tests in the two groups.

Table 18 presents the tests in rank order as measured by partial correlation with chronological age constant for the group of 50 children with Criterion I and for the group of 235 which includes the group of 50, with Criterion II. This table also shows *r*, PE and *i-k* for these tests.

TABLE 18

50 CHILDREN, CRITERION I					235 CHILDREN, CRITERION II				
Test	Rank	<i>r</i>	PE	<i>i-k</i>	Test	Rank	<i>r</i>	PE	<i>i-k</i>
Kingsbury ¹	1	.665	.053	.253	Otis	1	.854	.012	.480
Pintner-Cunningham	2	.607	.060	.205	Kingsbury	2	.836	.013	.451
Myers	3	.604	.061	.203	Myers	3	.818	.015	.425
Otis	4	.580	.063	.185	Pressey	4	.813	.015	.418
Dearborn	5	.521	.069	.146	Pintner-Cunningham	5	.810	.015	.414
Pressey	6	.463	.075	.114	Dearborn	6	.809	.015	.412
Haggerty	7	.440	.077	.102	Haggerty	7	.599	.028	.190

¹ This table should be read: The Kingsbury test ranks first by partial correlation (chronological age constant) with Criterion I, having a correlation of .665, with PE .053, and *i-k* .253, etc. The measure *i-k* indicates the improvement in estimate by a test over an estimate by chance, which in this test is .253.

The rank order of the tests for the two groups is as follows :

50 CHILDREN, CRITERION I	235 CHILDREN, CRITERION II
1. Kingsbury	1. Otis
2. Pintner-Cunningham	2. Kingsbury
3. Myers	3. Myers
4. Otis	4. Pressey
5. Dearborn	5. Pintner-Cunningham
6. Pressey	6. Dearborn
7. Haggerty	7. Haggerty

It will be noted that the correlations with Criterion II, which is the average of the GIQ's, are higher than are the correlations with Criterion I which is the IQ of the Stanford-Binet test. Because of this the IQ's themselves cannot be compared but the relative position of the tests in the groups can be compared. When this is done on the basis of the division upper four and lower three in the group of 50, only one test, the Pintner-Cunningham, is ranked as it is by the median deviation, though in the group of 50 both Otis and Myers hold Rank 4½ by the median deviation and both are in the upper four by Criterion I. As has been suggested before,¹ the difference in the ranking by these two criteria

¹ Page 34.

is not necessarily of much significance because of the difference in meaning of the two methods of measurement. In the lower three ranks one test, Haggerty, is ranked as it is by the median deviation. There is some significance in the fact that Haggerty is ranked lowest by both measures. When the group of 235 is compared with the ranking by median deviation not one of the tests agrees as to location in the upper four ranks and only the Haggerty test, which again ranks lowest, agrees as to location in the lower three. When these ranks by partial correlation are compared for the group of 235 with Criterion II with those for the group of 50 with Criterion I the division in halves will place the higher four in the upper half and the lower three in the lower half. On this basis three tests, Kingsbury, Myers and Otis, are in the upper half of the ranks by both criteria, and two, Dearborn and Haggerty, are by both criteria placed in the lower half with Haggerty taking the lowest rank.

When improvement of prediction as measured by 1-k of the correlations with Criterion I is considered it is found that Kingsbury test improves a chance prediction about 25% while Haggerty only gives an improvement of 10% with the others ranging between these two. In the 1-k for the correlations with Criterion II there is a much wider range of the improvement. Here Otis test gives an improvement of 48% over chance when the estimate made by that test is compared with the estimate which would result from the use of a combination of the seven tests weighted equally. On the same basis Haggerty test gives an improvement over chance of about 20%. In this case the improvement gained by the use of the Haggerty test is less than one-half of that gained by the Dearborn, which is the test which ranks next above it.

The correlation of Criterion I and Criterion II in this grade is .882. None of the tests in this grade reach a correlation with either criterion which closely approximates this, the highest correlation with Criterion II being .845 for Otis and the highest correlation with Criterion I being .665 for

Kingsbury. The meaning of these discrepancies is a matter for statistical determination.

The foregoing facts do not yield much conclusive evidence as to relative value of the tests but they do suggest certain uses. Measured by the median deviation, Pintner-Cunningham and Haggerty hold the same ranks in both groups, Pintner-Cunningham being second and Haggerty being seventh in rank; two other tests, Pressey and Dearborn, are in the upper three ranks by this measurement. This may indicate that if one desires to use a group test which will yield approximately the same individual measurement as a combination of the seven tests would yield one should use one of the three which hold the upper three ranks. Such a method of measurement leads to doubt as to the desirability of using the Haggerty test.

When ranked by correlation, the ranking by Criterion I and by Criterion II agrees in placing Kingsbury, Myers, and Otis tests in the upper half of the ranks and Dearborn and Haggerty tests in the lower half of the ranks. If what is desired is ranking the children for homogeneous groupings the Kingsbury, Myers or Otis tests may be used with the expectation that they will give a rating similar to that which would be given if the criteria were used, to a degree shown by their correlation with the criteria.

Table 19 presents in rank order the pairs of tests combined with equal weight in all possible combinations, for the group of 50 correlated with Criterion I. This gives 21 possible ranks for the combinations of the seven tests.

Table 20 shows another arrangement of the correlations of these mean IQ's with Criterion I. If this table is compared with the part of Table 18 which shows the correlations of the single tests for the group of 50 with Criterion I it is possible to determine the advantage to any test of making these combinations. Such a comparison shows that Myers and Otis tests are improved by every combination and also improve, by raising the correlation, every test with which they are combined; that the Dearborn, Haggerty and Pressey tests are also improved by every combination; but that Dear-

born lowers the correlation of Pintner-Cunningham alone; Haggerty gives the same correlation with Pintner-Cunning-

TABLE 19

SECOND GRADE—FIRST YEAR—50 CHILDREN. CORRELATION (CHRONOLOGICAL AGE CONSTANT) OF MEAN IQ'S ON PAIRS OF GROUP INTELLIGENCE TESTS WITH CRITERION I

Tests	Rank	r	PE	r-k
Kingsbury and Pintner-Cunningham ¹	1	.720	.046	.306
Kingsbury and Otis	2	.692	.050	.278
Kingsbury and Dearborn	3	.690	.050	.276
Kingsbury and Myers	4	.685	.051	.272
Kingsbury and Haggerty	5	.673	.052	.260
Myers and Pintner-Cunningham	6	.665	.053	.253
Haggerty and Myers	7	.660	.054	.249
Kingsbury and Pressey	8½	.659	.054	.248
Myers and Pressey	8½	.659	.054	.248
Dearborn and Myers	10	.658		
Myers and Otis	11	.647		
Otis and Pintner-Cunningham	12	.640		
Pintner-Cunningham and Pressey	13	.620		
Haggerty and Otis	14	.608		
Haggerty and Pintner-Cunningham	15	.607		
Dearborn and Otis	16	.595		
Dearborn and Pintner-Cunningham	17	.594	.062	.196
Otis and Pressey	18	.590	.062	.193
Haggerty and Pressey	19	.585	.063	.189
Dearborn and Pressey	20	.584	.063	.188
Dearborn and Haggerty	21	.582	.063	.186

¹ This table should be read: The mean IQ of the Kingsbury and Pintner-Cunningham tests ranks first when correlated with Criterion I (chronological age constant) and has a correlation of .720, with PE .046, and r-k .306, etc. The measure r-k indicates the improvement in estimate by a test over an estimate by chance, which in this test is .306.

TABLE 20

SECOND GRADE—FIRST YEAR—50 CHILDREN. CORRELATION (CHRONOLOGICAL AGE CONSTANT) OF MEAN IQ'S ON PAIRS OF GROUP INTELLIGENCE TESTS WITH CRITERION I

Tests	Hag.	Kg.	Myers	Otis	P-C	Pr.
Dearborn ¹	.582	.690	.658	.595	.594	.584
Haggerty		.673	.660	.608	.607	.585
Kingsbury			.685	.692	.720	.659
Myers				.647	.665	.659
Otis					.640	.590
Pintner-Cunningham						.620

¹ This table is to be read: The correlation between Criterion I and the mean of the Dearborn and the Haggerty IQ's is .582, etc.

ham as Pintner-Cunningham gives alone; Pressey lowers the single correlation of Kingsbury, by combination.

Aside from knowing with what tests others combine to advantage it is worth while to know the degree of advantage, that is, what the correlation obtained by the combination is and how much advantage is added by its use. For this it is necessary to consult Table 19. A survey leads to the conclusion that Kingsbury stands very high as a combining test. It is one of the pair in the first five ranks, and only with the Pressey test where it suffers by the combination does it fall as low as Rank 8 ($8\frac{1}{2}$ because the Myers-Pressey combination has the same correlation). In each of the first five pairs each test is helpful in the combination. One might conclude from this that if any single one of these tests is to be used it would be an advantage to add the Kingsbury test.

Another way of looking at it is to see how much improvement in estimate over chance estimate is gained by the combination. Taking the highest correlation of a Kingsbury combination, that is with Pintner-Cunningham, it will be seen that the advantage over chance is about 31%. For the Kingsbury test alone this advantage is about 25% (Table 18). There is not a great deal of difference here and it would be necessary to decide whether the increased degree of accuracy obtained was worth the extra expenditure of time, money and effort. The improvement for the Pintner-Cunningham test by the combination is much better than that gained for the Pintner-Cunningham test alone, a gain of 31% over 21%. In following down the table showing gain by 1-k it will be seen that the gain for the Otis test by combination with Kingsbury is the difference between 28% for the combination and 19% for the Otis alone. For the Dearborn test the increased gain by combination with Kingsbury is 28% less 15%, which indicates that the use of the Kingsbury test nearly doubles the value of the Dearborn test. If Dearborn was used it would seem desirable to use Kingsbury also. For the Myers test the gain by the combination with Kingsbury is 27% less 20%, or 7% greater efficiency than when Myers is used alone. For the Haggerty test a combination with Kingsbury gives increased efficiency as measured by 1-k of

26% less 10% or considerably more than double the improvement over chance which Haggerty gives alone. While other tests improve every test or almost every test with which they are combined no other combination ranks as high as these five with Kingsbury. It is interesting that while in the single correlations Myers outranks both Otis and Dearborn when these three tests are each combined with Kingsbury, the last two do not change their relative positions, but both outrank the Kingsbury-Myers combination. 1-k has been found for the first nine ranks and for the lowest five ranks, so that it is possible to compare the extremes of the range. This indicates a considerable advantage in the combinations at the upper extreme over the combinations at the lower extreme.

A study of the number of times each test appears in combination in the upper half of the range (11 ranks) and in the lower half of the range (10 ranks) is shown in the following summary:

Half	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Upper	2	2	6	6	2	2	2
Lower	4	4	0	0	4	4	4

It appears from this presentation that the Kingsbury and Myers tests give the best service in combination for second grade, as these two by every combination are located in the upper half of the range of ranks. The other five tests are distributed in exactly the same way, so that from this viewpoint there is nothing to choose among them. When the range is divided into thirds, with seven ranks in each third, somewhat more differentiation is made. This is shown by the following arrangement:

Half	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Upper	1	2	5	3	1	2	0
Middle	1	1	1	3	3	2	3
Lower	4	3	0	0	2	2	3

This suggests less value for the tests which in half of their combinations or more are in the lowest third of the range of ranks.

The range of ranks of the different tests in combination with others is as follows:

Test	Range in rank when combined with other tests
Kingsbury	1 - $8\frac{1}{2}$
Pintner-Cunningham	1 - 17
Otis	2 - 18
Dearborn	3 - 21
Myers	4 - 11
Haggerty	5 - 21
Pressey	$8\frac{1}{2}$ - 20

The lowest end of the ranks in each case indicates the first appearance of the test in combination. This gives some light on the ranking of the test. There is a greater discrepancy in the combination of the Dearborn test with others than for any other test. Haggerty and Pressey rank consistently low as is also indicated above by the plotting of the locations in thirds.

Summary

1. The Pintner-Cunningham test holds second rank when measured by the median deviation in both the group of 50 and the group of 235. This indicates that it stands relatively high as a valid measure of individual children when this validity is compared with the measurement by a combination of the seven tests weighted equally. The Haggerty test has the seventh rank for both groups which suggests lack of validity as measured by median deviation from Criterion II.

2. Pintner-Cunningham, Otis and Haggerty are ranked similarly by correlation with Criterion I for the group of 50 to their ranks by the median deviation. When correlation with Criterion II for the group of 235 is the basis of comparison only Haggerty test has a rank similar to its ranking by median deviation.

3. Kingsbury, Myers and Otis tests are ranked consistently in the upper half of the range of ranks when measured either by correlation with Criterion I for the group of 50 or by Criterion II for the group of 235. Dearborn and Haggerty are by the same criteria consistently ranked in the lower half of the range of ranks.

4. Measurement of the advantage to be gained in use of the test, by 1-k, shows a considerable difference between the upper and lower extremes of this range when either the group of 50 with Criterion I or the group of 235 with Criterion II is examined. The result for the Haggerty test for the group of 235 especially emphasizes its apparent inferiority as shown in most of its relationships in this grade, since the next test above shows more than double the advantage of the Haggerty test when measured by 1-k.

5. Kingsbury test is outstanding as a good test for combining with other tests of this group since it appears in combination in each pair in the first five ranks. Myers and Otis tests combine to advantage in an increased correlation with every other test. It appears to be undesirable to combine Dearborn and Pintner-Cunningham or Kingsbury and Pressey and without advantage to combine Haggerty and Pintner-Cunningham, since in each of the first two pairs one test has its correlation lowered by the combination and in the third pair no advantage is gained by one of the pair. In such a case it is uneconomical to use two tests where one of them will do as well or better than the two.

The statement has frequently been made that it is better to use two tests and average the results than to use only one test, for the sake of the greater validity of the measurement of the children. The above evidence indicates that this is true if the right tests are chosen for the combination, but that not *any two* tests will serve advantageously. It is apparent that if used indiscriminately one may harm the results rather than increase their validity.

6. The degree of improvement over a chance estimate of intelligence for correspondence with the Stanford-Binet IQ, as shown by 1-k, indicates a considerable difference in the desirability of the combinations which are at the upper end of the range and those at the lower end. In making a choice of tests for use, it is well not only to consider the increased advantage, but the absolute advantage. A combination of two tests may double the advantage which one of them has by itself and yet not give the advantage which another test or pair of tests will give.

7. When the range of ranks by partial correlation with Criterion I for the pairs of tests is divided into thirds, with seven ranks in each third, the advantages for single tests appear to be:

Kingsbury first, appearing five times in the highest third, once in the middle third and not at all in the lowest third; Myers, second, appearing three times in the highest third, and three times in the middle third; Pintner-Cunningham, third, appearing twice in each third; Haggerty, fourth, appearing twice in the highest third, once in the middle third, and three times in the lowest third; Otis, fifth, appearing once in the highest third, three times in the middle third and twice in the lowest third.

For Pressey and Dearborn it would be a toss-up for lowest place by this method of comparison. Pressey appears only three times in the lowest third and Dearborn four times, but Dearborn appears once in the highest third and Pressey not at all. This method of grouping gives to Haggerty the best rating which is given to it in this grade. Other things being equal, it would be better to make a combination for any test to be used which will throw its rank into the upper third of the ranks.

THIRD GRADE, FIRST YEAR

The results of the work of the third grade for 204 children are assembled in Tables 21 to 24.¹ These tables present the tests in rank order according to various criteria.

TABLE 21
RANK BY MEDIAN DEVIATION OF IQ FROM CRITERION II
50 CHILDREN 204 CHILDREN

Test	Rank	Median deviation	Test	Rank	Median deviation
Otis ¹	1	6.63	Dearborn	1	5.35
Myers	2	6.83	Otis	2	6.38
Dearborn	3½	7.00	Kingsbury	3	6.75
Kingsbury	3½	7.00	Myers	4	7.67
Haggerty	5	9.75	Haggerty	5	9.36

¹ This table is to be read: The Otis test has first rank with median deviation of 6.63, etc.

¹ For statement regarding original data see page 32.

Table 21 shows the rank order of the tests as indicated by the median deviation of the GIQ's from Criterion II for the group of 50 and for the group of 204 which contains the group of 50. This is taken from Table 5 in Chapter 3, where these data are compared to justify the use of the group of 50 as representative of the group of 204. As ranked by this criterion, tests for the two groups stand as follows:

50 CHILDREN	204 CHILDREN
1. Otis	1. Dearborn
2. Myers	2. Otis
3½. Dearborn	3. Kingsbury
3½. Kingsbury	4. Myers
5. Haggerty	5. Haggerty

Using the group of 50 as the basis of comparison it is found that the Otis test drops from 1 to 2 in rank, that Dearborn rises from 3½ to 1 in rank, that Myers drops from 2 to 4 in rank, that Kingsbury rises from 3½ to 3 in rank, and that Haggerty remains constant at rank 5. For the sake of making the comparison in both groups, both Dearborn and Kingsbury, which in the group of 50 have the same median deviation, will be called Rank 3. Then Dearborn, Otis and Kingsbury are in both groups in the first three ranks and Myers is the only test which is thrown out of the upper half (3 ranks) by the ranking in the larger group. This shows considerable consistency in the ranking.

TABLE 22

RANKS OF TESTS BY PARTIAL CORRELATION

50 CHILDREN, CRITERION I					204 CHILDREN, CRITERION II				
Test	Rank	r	PE	1-k	Test	Rank	r	PE	1-k
Haggerty ¹	1	.568	.065	.177	Dearborn	1	.959	.004	.717
Myers	2	.511	.070	.140	Kingsbury	2	.832	.015	.445
Otis	3	.468	.074	.116	Myers	3	.820	.016	.428
Kingsbury	4	.417	.079	.091	Otis	4	.645	.028	.236
Dearborn	5	.396	.080	.082	Haggerty	5	.627	.029	.221

¹ This table is to be read: The Haggerty test has first rank with r .568, PE .065, and $1-k$.177, etc. The measure $1-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .177.

Table 22 shows the ranks by partial correlation (chronological age constant) for the group of 50 with Criterion I and for the group of 204 with Criterion II.

In the group of 50 correlated with Criterion I there is a condition which is in marked contrast with the previous findings. The Haggerty test ranks first for this group on the basis of correlation with the Stanford-Binet IQ (Criterion I). A comparison of the ranking in this group with that by the median deviation shows that Myers and Otis are in the upper three ranks in agreement with the ranking by median deviation for the group of 50 but only Otis agrees with that ranking for the group of 204. The ranking for the lower half (2 ranks) agrees with neither of the groups in the median deviation ranking.

Comparison of the ranking of the group of 204 by correlation with Criterion II with the ranking by the median deviation shows Dearborn and Kingsbury in the upper half in agreement with the group of 50. In the lower half Haggerty agrees in its rank by correlation with Criterion II with its rank by the median deviation from Criterion II for both groups. The difference in the meaning of these two criteria should be recalled.¹

When the ranks of the group of 50 correlated with Criterion I are compared with the ranks of the group of 204 correlated with Criterion II it is found that Haggerty drops from first to last place, that only the Myers test is placed in the upper three by both rankings and that the Dearborn rises from last to first place when it is ranked by its relation to the average of the five tests instead of by its relation to the Stanford-Binet IQ. Myers and Otis each drop one step and Kingsbury rises two steps by this change in the basis of ranking. As in previous grades it will be noticed that the correlation with Criterion II is much higher than with Criterion I in which none of the tests are directly concerned.

Table 23 presents the tests in rank order as measured by the correlation of the mean GIQ's of pairs of tests with Criterion I for the group of 50. It also shows for each pair of tests *r*, PE, and 1-*k*. This last shows the increased accuracy over a chance estimate of intelligence when these estimates are compared with that obtained by the Stanford-Binet IQ.

¹ Page 34.

Table 24 presents these correlations of pairs of tests in an arrangement which makes easy a comparison of the amount gained by different tests when combined with others. A comparison of this table with the part of Table 22 which

TABLE 23

THIRD GRADE—FIRST YEAR—50 CHILDREN. CORRELATION (CHRONOLOGICAL AGE CONSTANT) OF MEAN IQ'S ON PAIRS OF GROUP INTELLIGENCE TESTS WITH CRITERION I

Test	Rank	r	PE	i-k
Haggerty and Myers ¹	1	.602	.061	.201
Haggerty and Otis	2	.598	.061	.198
Myers and Otis	3	.577	.064	.183
Haggerty and Kingsbury	4	.555	.066	.168
Dearborn and Otis	5½	.551	.066	.165
Dearborn and Myers	5½	.551	.066	.165
Dearborn and Haggerty	7	.547	.067	.163
Dearborn and Kingsbury	8	.516	.070	.143
Kingsbury and Myers	9	.513	.070	.142
Kingsbury and Otis	10	.482	.073	.124

¹ This table is to be read: The correlation between Criterion I and the mean of the Haggerty and Myers IQ's is .602, with PE of .061, and i-k .201, etc. The measure i-k indicates the improvement in estimate by a test over an estimate by chance, which in this test is .201.

TABLE 24

THIRD GRADE—FIRST YEAR—50 CHILDREN. CORRELATION (CHRONOLOGICAL AGE CONSTANT) OF MEAN IQ'S ON PAIRS OF GROUP INTELLIGENCE TESTS WITH CRITERION I

Tests	Haggerty	Kingsbury	Myers	Otis
Dearborn ¹	.547	.516	.551	.551
Haggerty		.555	.602	.598
Kingsbury			.513	.482
Myers				.577

¹ This table is to be read: The correlation between Criterion I and the mean of the Dearborn and the Haggerty IQ's is .547, etc.

shows for the group of 50 the rank of tests when correlated with Criterion I yields the following information:

Haggerty test drops in rank when combined with Dearborn or Kingsbury tests, but it is improved when combined with Myers or Otis tests.

Myers, Otis, Kingsbury and Dearborn tests are improved by every combination with others of these tests. Conversely Dearborn and Kingsbury are the only two tests which cause a drop in correlation by combination and they cause a drop

only in the Haggerty test. It would seem evident that these two tests should not be combined with Haggerty.

In Table 23 the pairs may be compared as to rank. There are 10 possible ranks for the five tests in pairs. A division into halves would place five pairs in each half. Because two pairs of tests have the same correlation and because these two are the fifth and six pairs it will be necessary to throw six pairs of tests into the upper half and four pairs into the lower half. This gives the combinations Haggerty and Myers, Haggerty and Otis, Myers and Otis, Haggerty and Kingsbury, Dearborn and Otis, Dearborn and Myers in the upper half. In accordance with the policy suggested of making no combination which lowers the correlation of one of the pair, the Haggerty and Kingsbury combination would be excluded from this group when considering the use of tests in combination. The comparison of this table with the part of Table 22 which gives the ranks for the group of 50 on the basis of correlation with Criterion I shows the amount gained by the combinations.

Haggerty and Myers have a correlation of .602 with 1-k .201. Haggerty alone has r .568 with 1-k .177. There would appear to be little advantage gained by using an extra test which increases the value of the test by so small an amount. Myers alone has r .511 and 1-k .140. The advantage to this test is more than to the Haggerty test, giving an increase in correlation of 9% and in increase in 1-k of 6%. These do not look large but in both these cases it should be noted that the correlations of the single tests with Criterion I for this group are low enough to make their predictive value questionable, and that any increase in these values ought to be regarded as advantageous.

In the second pair, Haggerty and Otis, Haggerty's improvement by combination is even less than in the first pair. Otis is relatively much more improved by the combination, advancing from r .468 to r .598 or 13% and the increase in predictive value as indicated by 1-k is 20% less 12%, or 8%. The entire table may be compared in the same way. Since all advantages gained are slight it would be necessary to decide in

any case whether they were worth the added expenditure of effort, time and money.

When the grouping by dividing the range of ranks into halves is made the location of separate tests is as follows:

Half	Dearborn	Haggerty	Kingsbury	Myers	Otis
Upper	2	3	1	3	3
Lower	2	1	3	1	1

This suggests the undesirability of Kingsbury as a combining test for this grade, with Dearborn rated next lowest in combining value, the other three being equal as far as this measure indicates.

Summary

1. The Dearborn, Kingsbury and Otis tests are in the upper ranks as measured by the median deviation of the GIQ's from Criterion II in both the group of 50 and the group of 204, and the Haggerty test is lowest in rank in both of these groups.

2. Myers and Otis tests agree in location in the upper three ranks when measured by correlation with Criterion I for the group of 50 and by median deviation from Criterion II for the same group. When this group as measured by correlation with Criterion I is compared with the group of 204 measured by median deviation only the Otis test holds the same rank.

3. Dearborn and Kingsbury tests are in the upper half of the ranks when measured by correlation with Criterion II for the group of 204 and this agrees with the ranking in both groups which are measured by the median deviation from Criterion II.

Myers is in the upper half as ranked by correlation with Criterion II for the group of 204 as it is when ranked for the group of 50 by median deviation from Criterion II.

Haggerty agrees in its ranking by correlation with Criterion II with its ranking for both groups by the median deviation from Criterion II.

4. Haggerty test which holds the highest rank when correlated with Criterion I for the group of 50 drops to the

lowest rank when correlated with Criterion II for the group of 204.

Myers is the only test which is in the upper half of the range of rankings by both of these criteria. Dearborn, which has the lowest rank according to correlation with Criterion I for the group of 50, has the highest rank when correlated with Criterion II for the group of 204. Kingsbury rises from fourth to second place when correlated with Criterion II for the group of 204. Myers and Otis maintain their relative positions but each drops one place in rank from that given by correlation with Criterion I, when measured by correlation with Criterion II for the group of 204.

5. All tests but Haggerty are improved as shown by increased correlation when combined with any other of the five tests. Haggerty loses by combination with the Dearborn and Kingsbury tests.

6. The advantage gained for any of these tests by combination in pairs is small. The justification for making such combinations would lie chiefly in the low correlations for the single tests and the possibility that any increase in these would be an advantage.

7. A comparison of these tests as to their appearance in combinations in the upper or lower half of the range of ranks indicates that the Haggerty, Myers and Otis tests are best for such combinations since each appears three times out of a possible four in the upper half; that Kingsbury is the poorest for combination since it appears three times out of a possible four in the lower half of the range; that Dearborn is equally divided and so not strongly indicated as a desirable combination.

SUMMARY OF FIRST YEAR'S DATA

Two bases of comparison of the tests have been used throughout the study of the first year's data. First, there is the comparison on the basis of ranks as given by the median deviation of the GIQ's from Criterion II for both the group of 50 and the larger group for each grade. Second, is the

comparison on the basis of rank as found by correlation of these tests:

1. For the group of 50 with Criterion I.
2. For the larger group with Criterion II.
3. For the tests combined in pairs with Criterion I.

A review of the summaries for each grade will help to determine whether there is any clear indication of the value of given tests throughout the range of their usefulness or whether, as far as this study shows, there are any specific grade preferences. To determine this each test will be traced through all the grades where it was given and the results compared. For convenience these will be considered in alphabetical order.

Cole-Vincent test was used in kindergarten and first grade. As measured by the median deviation it ranks first of the three tests used in the kindergarten in both the group of 50 and the group of 175. In the first grade it also ranks first in the group of 186 children and is second in the group of 50 children. Therefore, for the purpose of accuracy of measurement of individual children when the measurement of that accuracy is the degree of deviation from the measurement resulting from this combination of tests, Cole-Vincent ranks high. The combination with which comparison was made was three tests in the kindergarten and nine tests in the first grade.

When ranked by partial correlation with Criterion II for 175 children in the kindergarten Cole-Vincent is the middle one of the three tests given. It holds the same rank when measured by the Criterion I for the group of 50. In combination with each of the other tests it is one of each of the two lower pairs. In the first grade the correlation with Criterion I for the group of 50 children places Cole-Vincent in the upper half of the range of ranks (Rank 4), but correlation with Criterion II places it in the lower half (Rank 8). In the combinations of each test with others it appears only twice in the upper half of the ranks. This is in agreement with its ranking in the kindergarten. The relatively low ranking when measured by correlation with Criterion II might be explained on the basis of the make-up of Criterion II. This will not

explain its loss of prestige when combined with other tests and correlated with Criterion I which ranks it high when the correlation is with the single test. On the whole, Cole-Vincent presents more satisfactory results as compared with the other tests for these two grades when used alone. When combined, it is best combined with Pintner-Cunningham for the kindergarten and with Pressey or Myers for the first grade.

Dearborn test is the only test used in all four of the grades studied. Its results will therefore have an especial interest. First a survey of the results of its measurement by median deviation will be made. In the kindergarten Dearborn holds the middle rank of the three tests when measured by median deviation for both the group of 50 and the group of 175. In the first grade Dearborn holds first rank in the group of 50 and second rank in the group of 186, that is, it is well up in the upper third of the ranks. In the second grade Dearborn is one of the two tests tied for second rank in the group of 50 and holds first rank in the group of 235. In the third grade Dearborn ties with Kingsbury for third place (out of five) in the group of 50 and holds first rank in the group of 204. There is apparently enough correspondence in this rating to make it safe to say that for accurate rating of individuals on the basis of degree of agreement with this criterion the Dearborn test is satisfactory for the four grades studied.

The correlation of the single tests for the group of 50 with Criterion I shows the following condition. In the kindergarten the Dearborn test holds the lowest rank. In first grade Dearborn test holds the sixth rank out of nine, that is, it is in the lowest third of the range of ranks. In the second grade Dearborn holds fifth rank out of seven which is again low. In the third grade Dearborn has the lowest rank. So far as these results have significance they show that the Dearborn test seems unsatisfactory for children in these grades if the purpose is to rank the children as they would be ranked by the Stanford-Binet test.

When examined for the measurement in the large groups by Criterion II it is found that in the kindergarten, Dearborn test is ranked lowest as it is by Criterion I. In the first grade, Dearborn has a rank of six as it did in this grade when measured by Criterion I. In the second grade, Dearborn ranks sixth instead of fifth out of seven ranks as it did when ranked by Criterion I. In the third grade, measured by Criterion II for 204 children, Dearborn ranks first. This is inconsistent with all other rankings, even with its ranking in this grade with Criterion I where it ranked lowest. When measured then by either Criterion I or Criterion II through these four grades Dearborn ranks very low with the one exception of the third grade group of 204 correlated with Criterion II where it ranks first. As was suggested in an earlier connection it would be interesting to use some other basis for deriving Criterion II, for instance, the median GIQ, and see what difference it would make in the results. At present there is no explanation of this notable exception.

When examined as to its pairing with other tests it is seen that in the kindergarten it is in combination both first and last; that is, it combines well with Pintner-Cunningham and not nearly so well with Cole-Vincent. In first grade Dearborn appears only twice in the upper half, the first 18 of the ranks, and six times in the lower half. The combinations in the upper half are with Pressey and Otis tests. It seems to be an unsatisfactory combiner for the most part in this grade. In the second grade when the range of ranks is divided into thirds, with seven ranks in each third, Dearborn appears once in the upper third, once in middle third, and four times in the lower third. However, it does secure a place in the highest third where Pressey does not appear at all. In third grade Dearborn is equally divided between the upper and lower halves of the 10 ranks of the range. However, its appearances in the upper half are both in the two pairs which tied for fifth place and so are ranked $5\frac{1}{2}$. Strictly, it is questionable whether it should be rated in the upper half. The mass of evidence ranks the Dearborn test low as measured with the two criteria which were used for correlation.

This suggests that probably it is an inferior measure for classifying children into homogeneous groups on the basis of intelligence when compared with the grouping which would result when the children are classified either by the Stanford-Binet test or by a combination of all the tests used in the grade. The only exception to this has been noted above.

Detroit test was used only in first grade. It is ranked third by median deviation with both groups of children. That places it in the upper third of the range of ranks and suggests its suitability for valid measurement of individual children in harmony with this criterion.

When ranked alone by correlation it is first in rank when correlated with Criterion I for the 50 children but it is seventh in rank (out of nine ranks) when correlated with Criterion II for 186 children. This discrepancy again suggests experimenting with Criterion II to discover a possible improvement, and why the test changes from a position at the top of the upper third to the top of the lowest third. When measured by the satisfactoriness of its combination with other tests it holds a middle ground, being four times in each half. Its appearances in the upper half are: Rank 1 with Pressey; Rank 4 with Otis; Rank $5\frac{1}{2}$ with Pintner-Cunningham, and Rank 12 with Myers. These are all in the highest third and three are in the highest seventh of the ranks. The locations in the lower half are: Rank 20 with Dearborn; Rank 24 with Cole-Vincent; Rank 30 with Kingsbury, and Rank 31 with Haggerty. To a much more marked degree than any of the other tests, the Detroit is influenced by combining with any one of the other eight tests. This is evidenced by the wide range of ranks, from one to thirty-one; also by the rather even spread over the lower half in contrast with the concentration in the upper half. It seems to make considerable difference with what tests Detroit is combined. Care should be exercised not to combine it with any test which throws it into the lower half of the range since each of these tests lowers the correlation which Detroit has with Criterion I when used alone. Aside from its ranking in the group of 186 when measured by correlation with Criterion II, Detroit stands very well as a test for use in the first grade.

Haggerty was used in first, second, and third grades. As measured by median deviation in the first grade Haggerty ranks lowest for the group of 50 and eighth out of nine for the group of 186. In second grade Haggerty ranks seventh in both groups. In the third grade Haggerty ranks fifth in both groups. So far as the measurement by median deviation from Criterion II is concerned Haggerty seems an unsatisfactory test for accurate measurement of individuals.

When measured by correlation with Criterion I for the groups of 50 children, Haggerty in the first grade is ranked eighth out of nine possible ranks. In second grade Haggerty is seventh, the lowest rank; while in third grade there is the anomaly mentioned of Haggerty's ranking first. A possible explanation is that Haggerty is better suited to the third grade level than to either first or second grades. This would need further corroboration and evidence is not now available to prove it.

When ranked on the basis of correlation with Criterion II for the large groups, in the first grade Haggerty ranks ninth instead of eighth in the group of 50 correlated with Criterion I. In second grade Haggerty is seventh, in agreement with its ranking in the group of 50. In third grade it is fifth in contrast with its rank of first place for the group of 50 correlated with Criterion I.

When considered as to its usefulness in combination it appears that in the first grade it comes once in the upper half of the range of ranks and seven times in the lower half of the range. Its combination in the upper half is with Pressey, which appears to have in this grade particularly good combining value. In second grade Haggerty appears in combination in the upper third of the ranks (seven in number) twice, with Myers and Kingsbury in the middle third once, and in the lower third three times. While here it is better than either Dearborn or Pressey it still remains lower than all the others. In the third grade in combination Haggerty does as well as any other test, being equal to Myers and Otis tests in having three ranks in the upper half and one rank in the lower half of the range and is better than either Dearborn

or Kingsbury. While on the whole Haggerty makes a poor showing as a measure for ranking children on intelligence in agreement with either criterion, in the third grade there are notable exceptions when measured by correlation with Criterion I either alone or in combination; and some advantage also is apparent in second grade where it appears in the upper third, though at its lower end when combined with certain other tests.

Kingsbury test was used in first, second and third grades. In first grade when ranked by median deviation it is seventh out of nine ranks in both groups. In second grade it is sixth in rank out of seven in the group of 50, and fourth in the group of 235. In third grade it ties with Dearborn for third place in the group of 50 and holds third place in the group of 204. With the liberal interpretation of the "upper half," which has been used in this discussion, this throws Kingsbury in the lower half, as measured by this criterion in all cases, except for the group of 235 in the second grade, where it rises to fourth rank, and in both groups in the third grade. The preponderance of evidence indicates that Kingsbury test is not a very satisfactory measure for determining individual intelligence of the first and second grades in harmony with this criterion.

When measured by correlation with Criterion I for the groups of 50 the evidence is as follows: In first grade Kingsbury is ninth, or lowest in rank; in second grade Kingsbury is first in rank; while in the third grade Kingsbury is fourth out of five ranks. There is a definite lack of agreement here. When measured by Criterion II for the larger group in the first grade Kingsbury ranks four instead of nine as it did in the group of 50 with Criterion I. In the second grade Kingsbury ranks second instead of first as in the group of 50. In the third grade Kingsbury ranks second instead of fourth as in the group of 50. The second grade results are consistent for the two criteria. The third grade results more nearly agree with the second grade results for the group of 204 with Criterion II, while they agree better with the results of the first grade for the group of 50 with Criterion I. One

might conclude that Kingsbury is a satisfactory test for second grade but not especially to use for first grade and less satisfactory for third grade. This would be an unwarranted assumption without further evidence.

When grouped in pairs in the first grade, Kingsbury appears three times in the upper half and five times in the lower half of the range of ranks. Its appearances in the upper half are with Pressey, Pintner-Cunningham and Otis tests. These come within the ranks 8 to 13. The other five appearances come from ranks 25 to 32. Thus it would seem that Kingsbury is considerably influenced when combined with other tests. In second grade when the range of ranks is divided into thirds Kingsbury is found five times in the highest third, once in the middle third and not at all in the lowest third. This is a better record than that of any other test by this measurement. Myers, which approaches it most nearly, has no rank in the lowest third, but has three in each of the other thirds, which places it below Kingsbury. There is evidently some value in the Kingsbury in this grade, which invites further investigation. In the third grade Kingsbury appears once in the upper half and three times in the lower half of the range of ranks. This is the most inconsistent of the tests yet compared. It apparently has special usefulness in the second grade group. Further study is needed before positive statements can be made about this test.

Myers test was used in first, second and third grades. In first grade, as ranked by median deviation, it holds fifth rank in both groups. In second grade Myers holds fifth rank in the group of 235 and ties for fourth rank in the group of 50. This places it in the lower half in both groups. In third grade it holds second rank in the group of 50 and fourth rank in the group of 204. The only time this test achieves the upper half in the ranks by this criterion is in the third grade for the group of 50.

When measured by correlation with Criterion I for the group of 50, in first grade Myers ranks fifth, in second grade third, in third grade second. The regular rise in rank from grade to grade with this criterion suggests that Myers may

be more suitable for the higher than for the lower grades. When measured by Criterion II with the larger groups Myers stands fifth in the first grade as it did in the group of 50. In second grade Myers stands third as it did in the group of 50. In third grade Myers ranks third instead of second as it did in the group of 50. Here there is evidenced much the same advance with grade as in the ranking by Criterion I.

Considered as to combining value, it is found that in first grade Myers appears six times in the upper half of the range of ranks and twice in the lower half. This is better than any other test, except the Pressey, in this grade. In second grade Myers is next to Kingsbury in its combining value and above all other tests, appearing three times in the upper third of ranks and three times in the middle third. In third grade Myers appears three times in the upper half of the range of ranks and once in the lower half. It agrees with Haggerty and Otis tests in this division of the ranking. Myers is inconsistent in its ranking in different grades; but, if one may be allowed to say so, it is consistent in its inconsistency. It is a plausible assumption to make that the regular advancement from grade to grade may be attributable to better adaptation to the test as the children advance from grade to grade.

Otis test was used in first, second and third grade. In first grade, when measured by median deviation, Otis ties with Haggerty for the lowest rank in the group of 50 and is lowest in the group of 186. In second grade Otis ties with Myers for fourth place in the group of 50 and holds sixth place in the group of 235. In third grade Otis holds first place in the group of 50 and second place in the group of 204. There is somewhat the same situation here as with the Myers test, where there is a rather regular advancement through the grades.

Otis holds third rank in the first grade in the group of 50 measured by Criterion I, fourth rank in the second grade in the group of 50 measured by Criterion I, and third rank in third grade in the corresponding group measured by Criterion I. There is a rather consistent location of the test by

this criterion in the upper half of the range of each grade. When measured by Criterion II for the larger groups, Otis holds first rank in first grade, in second grade it holds first rank, and in third grade it holds fourth rank. These correlations confirm the ranks by Criterion I except in the last instance, that of the third grade group of 204 ranked by Criterion II.

In combination in pairs it is shown that in the first grade Otis appears five times in the upper half of the range of ranks and three times in the lower half. This is the same rating that Pintner-Cunningham has and these two are next to Pressey and Myers in the frequency of their combinations in the upper half of the ranks. In second grade Otis does not stand quite so well, holding one rank in the highest third, three ranks in the middle third, and two ranks in the lowest third of the range. In the third grade Otis holds three places in the upper half of the range and only one in the lower half. That is, with the two exceptions noted, when ranked by correlation with the criteria alone or with Criterion I when paired with other tests, Otis stands well in relation to other tests in this group.

Pintner-Cunningham test was used in the kindergarten, first and second grades. In kindergarten, when ranked by median deviation, the Pintner-Cunningham stands lowest in both groups. In first grade it stands sixth in the group of 50 and fourth in the group of 186. In second grade it ties with Dearborn for second place in the group of 50 and holds second place in the group of 235. As measured by this criterion Pintner-Cunningham seems to be increasingly satisfactory from kindergarten through second grade.

When measured by correlation with Criterion I Pintner-Cunningham stands first in the kindergarten, second in the first grade and second in the second grade. It seems reasonable to conclude from this evidence that Pintner-Cunningham is relatively satisfactory for ranking children and conforms with their ranking by the Stanford-Binet test. When ranked with Criterion II for the larger groups of children, Pintner-Cunningham holds first place in the kindergarten as it did

in the group of 50. It has second place in the first grade as it did in the group of 50. It drops to fifth place in the second grade instead of second place as in the group of 50. This is one more instance of inconsistency when the ranking is by correlation with Criterion II. The other rankings suggest that the test is satisfactory for ranking the children in these grades.

Pintner-Cunningham is the most satisfactory of the tests for combination in pairs in the kindergarten as it appears in both the first and the second pairs. In first grade Pintner-Cunningham appears five times in the upper half of the range and three times in the lower half. This makes it equivalent to Otis for combinations and next to Pressey and Myers. In second grade Pintner-Cunningham appears twice in each third of the range. It is here less satisfactory than Kingsbury or Myers and more satisfactory than the other four tests, Dearborn, Haggerty, Otis, Pressey. The weight of evidence seems to be in favor of Pintner-Cunningham as a good test for ranking children, with some little question in second grade for purposes of combination. However, it ranks in the upper half of the tests even on this rating.

Pressey test was used in first and second grades. In first grade in the group of 50 it is ranked fourth by the median deviation and in the group of 186 it is ranked sixth. In the second grade it is ranked first in the group of 50 and third in the group of 235. In only one instance did it rank lower than the upper half of the range of ranks. By this criterion, then, it may be regarded as relatively good for the purpose which is intended by this criterion.

In first grade Pressey ranks seventh when measured by correlation with Criterion I for the group of 50. For the corresponding group in the second grade it ranks sixth. This suggests that used alone it is relatively an unsatisfactory measure for determining rank order of children in intelligence as obtained by Stanford-Binet test. When ranked by Criterion II for the larger groups Pressey ranks third in the first grade and fourth in the second grade. The correspondence seems to be better with the average of all the tests used than with the Stanford-Binet tests.

Combined in pairs, in first grade Pressey excels all other tests, appearing eight times in the upper half of the range of ranks. It comes eight times in the first 13 ranks and seven of these are within the first nine ranks. It is the one test which, in this grade, improves every other test and is improved by every other test. In second grade it is not so satisfactory, not appearing at all in the upper third of the ranks and appearing three times in each of the other two thirds. It is lower than any other test except Dearborn in this grouping. There is some indication here of the superiority of this test for use in the first grade in combination with other tests, even though it appears to be relatively inefficient when used alone. If a ranking is desired, which would be in harmony with that obtained by all the tests used for the grade, Pressey would probably be satisfactory for either grade. It is evidently not so good for combination in second grade as in the first grade.

CHAPTER 5

THE SECOND YEAR'S STUDY

After a year's interval the tests given to these grades were repeated for half the children in each grade who had taken them the first year. An ideal method of choosing the children for the repetition would have been to select on the same basis of choice as that used the first year. This was impossible because of the change in the school population. It was planned to keep the proportions for each school the same as the first year, taking one-half of the children tested the first year from each grade in each school. Even this was not always possible. In one school, for one grade, less than one-half the children tested the first year were found in the school the second year. All of these were tested, but it was necessary to make up the deficiency from the other schools. The first plan was to repeat the tests for only 50 children. The experience the first year with absences and the necessity of making up tests led to a change of plan. The procedure followed was to retest 100 children from each grade and make up no tests that were lost. This proved to be satisfactory as it gave more than 90 children in each grade who finished all the tests. Another difficulty found in the repetition of the tests was the fact that in many cases the children who were together the first year were in different rooms and sometimes in different grades the second year. For example, in one first grade class where it was necessary to use all the children remaining in the school a few were still in the first grade, one was in the third grade and the rest were divided between the 2B and 2A classes. This made it impossible to classify by grade the second year. The class mentioned above would naturally have been called the second year 2A, but it was made up of children in 1A, 2B, 2A, and 3A, and could not be truly called 2A. The method adopted was to call each grade by its first year's standing and add the term "retest." There is another reason

why this is a desirable method. For purposes of comparison the testing in the second year involved the giving of the same tests to each group of children which they had had the first year. It should be held in mind then that "Kindergarten Retest" means the children who were tested in the kindergarten the first year and retested the second year by the same three tests which were given to them the first year. "First Grade Retest" means those children tested in the first grade the first year by nine tests, who were retested the second year by the same nine tests. "Second Grade Retest" means those children tested in the second grade the first year by seven tests, who were retested the second year by the same seven tests. "Third Grade Retest" means those children tested the first year by five tests, who were retested the second year by the same five tests.

This method of retesting makes possible a number of interesting comparisons which might give light on the subject of growth from year to year through these grades. This question, however, is not a part of the writer's problem of the choice of tests and therefore is not attacked in the present report. It is planned to make a separate study of this later.

There was no attempt the second year to check for practice effect of tests, and they were given to all groups in alphabetical order except that in some cases Pressey test, which is a very short test, was sandwiched in where there would not have been time for a long test.

The group testing program for the second year occupied the time from September 28 to December 1, 1925. Not all the 220 children who had taken the Stanford-Binet test the first year were available for retesting, so that it was necessary to give more Stanford-Binet tests in order to study 50 children from each grade who had all the group tests and also the Stanford-Binet test. The number of Stanford-Binet tests which had to be given the second year were as follows:

Kindergarten	First	Second	Third grades	Total
11	14	6	12	43

All the tests the second year were given and scored by the writer. All tests were scored and the results returned to the

schools by the close of the first half-year. The data for each grade will appear in the order of presentation for the first year's work.

KINDERGARTEN, RETEST

Table 25 shows the data for kindergarten retest¹ assembled in such a way as to show the rank order of the tests for the group of 50 with Criterion I and for the group of 96 with Criterion II.

TABLE 25
KINDERGARTEN RETEST—SECOND YEAR. RANKS OF TESTS BY
PARTIAL CORRELATION

50 CHILDREN, CRITERION I					96 CHILDREN, CRITERION II				
Test	Rank	r	PE	r-k	Test	Rank	r	PE	r-k
Dn. ¹	1	.754	.041	.343	P-C	1	.896	.014	.556
C-V	2	.728	.045	.314	Dn.	2	.884	.015	.533
P-C	3	.636	.057	.228	C-V	3	.876	.016	.518

¹ This table should be read: The Dearborn test is first in rank by partial correlation with r .754, PE .041, $r-k$.343, etc. The measure $r-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .343.

This rank order for the group of 50 is Dearborn, Cole-Vincent, Pintner-Cunningham; and for the group of 96, Pintner-Cunningham, Dearborn, Cole-Vincent. While the Dearborn and Cole-Vincent tests do not change their relative positions by the two rankings they each drop down one step to make way for Pintner-Cunningham, which holds last place for the group of 50 and first place for the group of 96. As is to be expected the correlations are higher with Criterion II, of which each test furnishes one-third, than with Criterion I, which is not directly related to any of the three group tests. Considering this there is surprisingly little difference in the correlations of Dearborn and Cole-Vincent with the two criteria. It is worth while to notice also, what would be expected, that the range of the correlations with Criterion II is much less (only 2%) than it is with Criterion I (11%). Other things being equal, Criterion I would generally be accepted as the better criterion.

¹ For statement regarding original data see page 32.

The improvement over estimate by chance is found to be as follows: When compared with the estimate which is obtained by the use of the Stanford-Binet test Dearborn gains about 34% over a chance estimate, Cole-Vincent, 31%, and Pintner-Cunningham, 23%. These all indicate rather significant advancement by the use of the test instead of mere guessing, with quite an advantage of Dearborn over Pintner-Cunningham for this purpose. When compared with the estimate of intelligence which is obtained by the use of the three tests for the group of 96, the gain over chance estimate is for Pintner-Cunningham, 56%, for Dearborn, 53%, and for Cole-Vincent, 52%. Again it will be seen here that there is relatively little difference between the tests as compared by Criterion II. Only a gain of 4% would accrue from the use of Pintner-Cunningham over what would be gained by the use of Cole-Vincent.

A comparison of these results with corresponding results for the first year shows the rank orders to be as follows:

Group of 50		Whole group	
First year	Second year	First year, 175	Second year, 96
1. Pintner-Cunningham	1. Dearborn	1. Pintner-Cunningham	1. Pintner-Cunningham
2. Cole-Vincent	2. Cole-Vincent	2. Cole-Vincent	2. Dearborn
3. Dearborn	3. Pintner-Cunningham	3. Dearborn	3. Cole-Vincent

It is seen that Pintner-Cunningham test stands three times out of four in first place, the first year with Criterion I and both first and second years with Criterion II. Its other rank is third, the second year with Criterion I. Cole-Vincent stands second in rank three times, both years with Criterion I and the first year with Criterion II. Its other rank is third, the second year with Criterion II.

Dearborn test stands third in rank twice, by both criteria the first year, in the first rank once by Criterion I the second year and in second rank once by Criterion II the second year.

Summary

This cumulative evidence indicates that these three tests would rank Pintner-Cunningham, Cole-Vincent, Dearborn.

This is the rank given by both criteria the first year. The results of the second year are not sufficiently at variance with this to change the order when the data are massed.

FIRST GRADE, RETEST

In Table 26 the data for the First Grade Retest group are given.¹ The co-efficients of correlation together with their

TABLE 26
DATA FOR FIRST GRADE—SECOND YEAR. RANKS OF TESTS BY
PARTIAL CORRELATION

50 CHILDREN, CRITERION I					91 CHILDREN, CRITERION II				
Test	Rank	r	PE	1-k	Test	Rank	r	PE	1-k
Dn. ¹	1	.786	.036	.382	Kg.	1	.903	.013	.570
P-C	2	.754	.041	.343	Pr.	2	.887	.015	.538
C-V	3	.739	.043	.326	Otis	3	.860	.018	.490
Pr.	4	.718	.046	.304	P-C	4	.843	.021	.462
Kg.	5	.707	.048	.293	Dn.	5	.835	.021	.450
Otis	6	.686	.051	.272	Myer	6	.794	.026	.392
Dt.	7	.644	.056	.235	C-V	7	.793	.026	.391
Myer	8	.617	.059	.213	Hag.	8	.744	.032	.332
Hag.	9	.515	.070	.143	Dt.	9	.731	.033	.318

¹ This table should be read: The Dearborn test ranks first with r .786, PE .036 and $1-k$.382, etc. The measure $1-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .382.

PE and $1-k$, are arranged so that the tests are shown in their rank order as measured by the two criteria used. $1-k$ for this table was obtained by reading k from Miner's Tables for $\sqrt{1-r^2}$ and making the subtraction for $1-k$.

The ranks of the tests by the two criteria are as follows:

CRITERION I

1. Dearborn
2. Pintner-Cunningham
3. Cole-Vincent
4. Pressey
5. Kingsbury
6. Otis
7. Detroit
8. Myers
9. Haggerty

CRITERION II

1. Kingsbury
2. Pressey
3. Otis
4. Pintner-Cunningham
5. Dearborn
6. Myers
7. Cole-Vincent
8. Haggerty
9. Detroit

There is some variation in the ranking by the two criteria. The amount of displacement will be seen if the group of 50

¹ For statement regarding original data see page 32.

is taken as the basis for comparison and the rank in each group is indicated with the difference in these ranks.

Test	Criterion I	Criterion II	Change in rank
Dearborn	1	5	-4
Pintner-Cunningham	2	4	-2
Cole-Vincent	3	7	-4
Pressey	4	2	+2
Kingsbury	5	1	+4
Otis	6	3	+3
Detroit	7	9	-2
Myers	8	6	+2
Haggerty	9	8	+1

If the range of ranks is divided into halves, five in the upper half and four in the lower half, it is found that Dearborn, Pintner-Cunningham, Pressey and Kingsbury tests are in the upper half by both measures and that Detroit, Myers and Haggerty tests are in the lower half by both criteria. Only Otis and Cole-Vincent tests shift from one half to the other.

When the range is divided into thirds, with three ranks in each third, there is no test which is in the highest third by both measures, nor is there any in the middle third. Two tests, Detroit and Haggerty, are placed in the lowest third by both criteria. It is an interesting situation where there is a sufficient degree of correspondence in the ranking by the two criteria so that only two tests are thrown from one half to the other and yet a large enough discrepancy in the rankings so that only two tests maintain their positions in the same third of the ranks by both criteria. There are three tests which change their ranks by four steps, one of these, Cole-Vincent, moving to the lower half as a result. The other two remain in the upper half but exchange places, Dearborn being first as ranked by Criterion I and fifth as ranked by Criterion II and Kingsbury being ranked fifth by Criterion I and first by Criterion II. Otis test is the only one that changes its rank three places and it is thrown from the lower to the upper half of the range by this change.

Four tests change their rank by two steps. Pintner-Cunningham and Pressey exchange places and both remain in the upper half of the ranks. Detroit and Myers each change

two steps, Detroit downward and Myers upward, but both remain in the lower half of the range.

Haggerty is ranked one step higher by Criterion II than by Criterion I.

As previously stated it is natural that correlations with Criterion II are higher than with Criterion I. The range of correlations with Criterion II is from r .903 to r .731 or .172 points. The range of correlations with Criterion I is r .786 to r .515 or .271 points, 10% greater range than with Criterion II.

When measured by the degree of correspondence to these criteria as shown by the advantage gained over chance by the use of a test, it will be found that the correlation with Criterion II gives a range of r -k from 57% to 32% or 25%. The correlations with Criterion I give a range of r -k from 38% to 14% or 24%. The comparison of the range of coefficients and the range of r -k indicates a closer grouping of the tests about Criterion II than the grouping about Criterion I for the same tests. This may be explained by the fact that in Criterion II each of the tests participates to the extent of one-ninth, while none of the tests participates directly in Criterion I.

A comparison of the rankings found in the use of these tests with the rankings of the same tests for the same children the first year may be interesting. This is as follows:

TABLE 27
RANKING OF TESTS GIVEN TO FIRST GRADE

Group of 50				Whole group			
First year		Second year		First year, 186		Second year, 91	
1. Detroit		1. Dearborn		1. Otis		1. Kingsbury	
2. Pintner-		2. Pintner-		2. Pintner-		2. Pressey	
	Cunningham		Cunningham		Cunningham	3. Otis	
3. Otis		3. Cole-Vincent		3. Pressey		4. Pintner-	
4. Cole-Vincent		4. Pressey		4. Kingsbury			Cunningham
5. Myers		5. Kingsbury		5. Myers		5. Dearborn	
6. Dearborn		6. Otis		6. Dearborn		6. Myers	
7. Pressey		7. Detroit		7. Detroit		7. Cole-Vincent	
8. Haggerty		8. Myers		8. Cole-Vincent		8. Haggerty	
9. Kingsbury		9. Haggerty		9. Haggerty		9. Detroit	

When these ranges are divided on the basis of halves, which was found to yield the best correspondence in the second year's work, the following condition is found:

Pintner-Cunningham appears in each of the four arrangements in the upper half of the range of ranks.

Otis appears three times in the upper range of ranks, the exception being the second year measured by Criterion I where it falls just below the upper half, having a rank of six.

Cole-Vincent appears twice in the upper half, as ranked by Criterion I both years. This shows an interesting correspondence with Stanford-Binet test which was made the basis for the standardization of the Cole-Vincent test as contrasted with a lack of harmony with the average of all the tests to which Cole-Vincent itself contributes one-ninth.

Myers appears twice in each half, being in the upper half by both measures the first year. In the ranking by Criterion II the second year it misses the upper half by one step, ranking sixth.

Dearborn appears twice in each half. The first year it just misses the upper half, ranking sixth in both groups and the second year it is placed in the upper half by both criteria.

Pressey ranks in the upper half three times, being in the lower half only in the first year when ranked by Criterion I. It will be recalled that this is the test which, as ranked by Criterion I the first year, showed exceptionally good combining value with the other tests. Haggerty ranks in the lower half in all four groups.

Kingsbury ranks in the upper half except as ranked by Criterion I the first year.

The frequency of the tests in the upper and lower halves in the four groupings is shown below.

Half	C-V	Dn.	Dt.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Upper	2	2	1	0	3	2	3	4	3
Lower	2	2	3	4	1	2	1	0	1

Summary

In the second year's work there is obviously a discrepancy in the ranking of the tests as indicated by the fact that only

two of the tests, Detroit and Haggerty, fall into the same third of the range of ranks as measured by Criterion I and Criterion II. However, these discrepancies are small enough in amount so that only two of the tests, Cole-Vincent and Otis, are not in the same half of the range by the two criteria. Dearborn, Pintner-Cunningham, Pressey and Kingsbury hold ranks in the upper half of the range when ranked by each of the two criteria, and Detroit, Myers and Haggerty are ranked consistently in the lower half.

In the first year's work it was found that Pintner-Cunningham, Otis and Myers ranked consistently in the upper half by the two criteria and that Dearborn and Haggerty ranked consistently in the lower half. The only two complete agreements for the two years are for Pintner-Cunningham in the upper half and for Haggerty in the lower half. This would at least seem to differentiate these two tests as to satisfactoriness at this level.

When the rankings for the two years are compared on the basis of the division of the range into halves the situation is found to be as follows: Pintner-Cunningham ranks first, always being in the upper half. Kingsbury, Otis and Pressey rank second, having three ranks in the upper half and one rank in the lower half. This disposes of four ranks.

Cole-Vincent, Dearborn and Myers rank fifth, each being equally divided between the halves. This takes seven ranks.

Detroit ranks eighth, appearing once in the upper half and three times in the lower half of the range.

Haggerty ranks ninth, appearing only in the lower half of the range.

If the range is divided into thirds and the ranking for the two years compared the following arrangement is shown:

Third	C-V	Dn.	Dt.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Highest	1	1	1	0	1	0	3	3	2
Middle	1	3	0	0	2	3	1	1	1
Lowest	2	0	3	4	1	1	0	0	1

Here Pintner-Cunningham and Otis tie for first rank with three ranks in the highest third and one in the middle third. Pressey ranks third with two ranks in the highest third, and

one in each of the other thirds. Four tests have each one rank in the highest third. Of these Dearborn stands highest because of having three ranks in the middle third and none in the lowest third. This gives it fourth rank. Kingsbury stands fifth with one rank in the highest third, two in the middle third and one in the lowest third. Cole-Vincent stands sixth with one each in the highest and middle thirds and two in the lowest third. Detroit stands seventh with one place in the highest third, none in the middle third and three in the lowest third. Myers stands eighth with none in the highest third, three in the middle third and one in the lowest third. On the rankings just shown it might be argued that Myers is superior to Detroit in having three rankings in the middle third. Haggerty ranks ninth with all its rankings in the lowest third.

The rankings of the tests on the basis of division of the range into halves and into thirds may be compared.

Division :

HALVES		THIRDS	
Test	Rank	Test	Rank
1. Pintner-Cunningham	1	1. Pintner-Cunningham	1½
2. Pressey	3	2. Otis	1½
3. Otis	3	3. Pressey	3
4. Kingsbury	3	4. Dearborn	4
5. Cole-Vincent	6	5. Kingsbury	5
6. Dearborn	6	6. Cole-Vincent	6
7. Myers	6	7. Detroit	7
8. Detroit	8	8. Myers	8
9. Haggerty	9	9. Haggerty	9

Six of the nine tests are given the same rank order by both treatments: Pintner-Cunningham, Pressey, Dearborn, Kingsbury, Cole-Vincent and Haggerty. Detroit and Myers merely change places, each changing one step, while Otis changes 1½ steps. It would seem that there is here rather consistent rating of these tests.

Since this class, the second year, would normally be second grade instead of first grade, a comparison might be made between the results of the second year's work and those of the second grade for the first year. This will be given later.

SECOND GRADE, RETEST

The data for the Second Grade Retest group are given in Table 28.¹ These show the tests in rank order as measured by Criterion I for the group of 50 and by Criterion II for the group of 97. The r , PE and $1-k$ are shown for each test. $1-k$ was found from k , which was read from Miner's Tables $\sqrt{1-r^2}$.

TABLE 28
SECOND GRADE RETEST—SECOND YEAR—97 CHILDREN. RANKS OF
TESTS BY PARTIAL CORRELATION

50 CHILDREN, CRITERION I					97 CHILDREN, CRITERION II				
Test	Rank	r	PE	$1-k$	Test	Rank	r	PE	$1-k$
Hag. ¹	1	.697	.049	.283	Kg.	1	.832	.021	.445
Otis	2	.556	.066	.169	P-C	2	.805	.024	.407
Myer	3	.504	.071	.136	Dn.	3	.801	.025	.401
Dn.	4	.482	.073	.124	Myer	4	.788	.026	.384
P-C	5	.472	.074	.118	Pr.	5	.777	.027	.370
Pr.	6	.469	.074	.117	Hag.	6	.764	.029	.355
Kg.	7	.355	.083	.065	Otis	7	.760	.029	.350

¹ This table should be read: Haggerty test ranks first with r .697, PE .049, and $1-k$.283, etc. The measure $1-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .283.

An examination of Table 28 will show some differences in the ranks of the tests by the two criteria.

Haggerty ranks first for the group of 50 by Criterion I, which is in itself surprising. It may be recalled that the only high rank given to Haggerty test in the first year was in the third grade with Criterion I. Since this second grade retest class would normally be called third grade there is an interesting suggestion here. This will be examined later. By Criterion II for the 97 children Haggerty is ranked sixth.

Otis ranks second by Criterion I but seventh by Criterion II.

Myers slips from third rank to fourth in the group of 97 correlated with Criterion II.

Dearborn moves from fourth place by Criterion I to third place by Criterion II.

Pintner-Cunningham rises from fifth rank by Criterion I to second place by Criterion II.

Pressey rises one step from sixth rank by Criterion I to fifth rank by Criterion II.

¹ For statement regarding original data see page 32.

Kingsbury rises from lowest rank by Criterion I to highest rank by Criterion II. The change in rank and amount and direction of the change is shown in the following list:

Test	Criterion I	Criterion II	Change in rank
Haggerty	1	6	-5
Otis	2	7	-5
Myers	3	4	-1
Dearborn	4	3	+1
Pintner-Cunningham	5	2	+3
Pressey	6	5	+1
Kingsbury	7	1	+6

This gives a wider range in the ranks by the two criteria than was found in the earlier grades. If a comparison is made on the basis of division of the range into upper half (4 ranks) and lower half (3 ranks) it is found that Myers and Dearborn are ranked in the upper half by both criteria and that Pressey test is the only one ranked by both in the lower half.

Haggerty and Otis are ranked in the upper half by Criterion I but are shifted to the lower half by Criterion II, though they hold the same relative positions with reference to each other.

Pintner-Cunningham and Kingsbury, ranked in the lower half by Criterion I, are shifted to the upper half by Criterion II and Kingsbury advances above Pintner-Cunningham, which was above it in the ranking by Criterion I.

The range of ranks by Criterion II is from $r .832$ to $r .760$ or $.072$ points, and by Criterion I is from $r .697$ to $r .355$ or $.342$ points, nearly five times the range by Criterion II. This agrees with the greater concentration found of the correlations about Criterion II in the first grade.

When the advantages of the tests as measured by r - k are considered in relation to the two criteria it is found that the correlations with Criterion I yield r - k ranging from 28% to 7%, a difference of 21%. The correlations with Criterion II yield a r - k range from 45% to 35%, a difference of 10%. When a similarity of ranking to that obtained by Stanford-Binet is desired it seems important to choose tests from the upper end of the range. It is interesting that more

than half of the difference between the extremes of the range of 1-k here occurs between the first and the second ranks where the 1-k difference is 11%.

The data in this table may be compared with the corresponding data for the first year with the idea of discovering confirmation or contradiction of the evidence found there.

TABLE 29
RANKING OF TESTS FOR SECOND GRADE AND FOR SECOND GRADE RETEST

50 children				Whole group			
First year		Second year		First year, 235		Second year, 97	
1. Kingsbury		1. Haggerty		1. Otis		1. Kingsbury	
2. Pintner-		2. Otis		2. Kingsbury		2. Pintner-	
Cunningham		3. Myers		3. Myers		Cunningham	
3. Myers		4. Dearborn		4. Pressey		3. Dearborn	
4. Otis		5. Pintner-		5. Pintner-		4. Myers	
5. Dearborn		Cunningham		Cunningham		5. Pressey	
6. Pressey		6. Pressey		6. Dearborn		6. Haggerty	
7. Haggerty		7. Kingsbury		7. Haggerty		7. Otis	

When these groups are divided into the upper half (4 ranks) and the lower half (3 ranks) the following distribution is found in these halves:

Half	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
Upper	2	1	3	4	3	2	1
Lower	2	3	1	0	1	2	3

It is noteworthy that there is no test which does not in some group occur in the upper half. The ranking on the basis of this grouping would be:

Myers, first, coming four times in the upper half.

Kingsbury and Otis second, coming three times in the upper half and once in the lower half.

Pintner-Cunningham and Dearborn fourth, coming twice in each half.

Pressey and Haggerty sixth rank, coming once in the upper half and three times in the lower half.

In an attempt to find a clearer discrimination the range was divided into three parts, having three ranks in the highest and two ranks in each of the others. This gave the following arrangement:

Part	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
High	1	1	3	3	2	2	0
Middle	2	0	0	1	1	2	2
Low	1	3	1	0	1	0	2

This arrangement would place Myers first and Kingsbury second because while each has the three ranks in the high group, Myers' other rank is in the middle group and Kingsbury's is in the lowest group. On the same basis Pintner-Cunningham is given third rank with two places in the high group and two places in the middle group, while Otis comes fourth with two ranks in the high group but one in each of the other divisions.

Dearborn then would have fifth rank with one place in the high group, two in the middle group and one in the low group, while Haggerty takes sixth place with one in the high group, none in the middle group and three in the lower group. There is opportunity for controversy as to the relative position of Haggerty and Pressey. Pressey is given seventh rank with no rating in the high group and two ranks in the low group. It is an open question whether two rankings in each of the two lower groups should place it above Haggerty, which has all but one ranking in the low group when that one rank is in the high group. The matter was determined here by the fact that the one rank in the high group made it natural to compare the Haggerty test with the Dearborn, which also had one rank in the high group.

Summary

From the data for the second grade retest, in the second year it appears that the relatively consistent rankings on the basis of division of the range into halves are Dearborn and Myers tests ranked by both criteria in the upper half and Pressey test ranked by both in the lower half of the range. Aside from these three tests Pintner-Cunningham test is the only test which has as little as three places difference in rank by the two criteria.

When these data were compared with corresponding data for the first year the range was divided in two different ways: 1. Into upper half and lower half; 2. into highest three ranks, middle two ranks and lowest two ranks. The correspondence of the ranks given tests by these two groupings is shown by the following arrangement:

UPPER FOUR, LOWER THREE		HIGHEST THREE, MIDDLE TWO, LOWEST TWO	
Test	Rank	Test	Rank
Myers	1	Myers	1
Kingsbury	2 $\frac{1}{2}$	Kingsbury	2
Otis	2 $\frac{1}{2}$	Pintner-Cunningham	3
Pintner-Cunningham	4 $\frac{1}{2}$	Otis	4
Dearborn	4 $\frac{1}{2}$	Dearborn	5
Pressey	6 $\frac{1}{2}$	Haggerty	6
Haggerty	6 $\frac{1}{2}$	Pressey	7

A comparison shows that Myers ranks first by both groupings and Kingsbury second; that Otis, Pintner-Cunningham and Dearborn tests hold the next three ranks in both arrangements with a shift which exchanges the rank positions of Pintner-Cunningham and Otis; that Pressey and Haggerty hold the lowest places in each grouping.

While there is no evidence of complete correspondence in these ratings, it is interesting that as a whole the order of the tests is practically the same for the two groupings of ranks where all corresponding ranks for the two years are included. A comparison of these ranks with those obtained from the corresponding data for the first year shows that Myers, Kingsbury and Otis tests held ranks in the upper half for the first year as they do in these composite ranks and that Dearborn and Haggerty held positions in the lower half the first year as they do in these composite rankings. The two tests which are at variance are the Pintner-Cunningham and the Pressey. The Pintner-Cunningham test the first year was dropped just below the upper half to fifth place by Criterion II, and the Pressey test gained fourth rank the first year by Criterion II, which gave it one place in the upper half of the range of ranks.

Considering the many variations within these groups, it is encouraging to note their small size and the general correspondence of rating of the tests by a combination of these ratings in the two methods of dividing the range.

The comparison of these data with corresponding results from the third grade work of the first year will be made later.

THIRD GRADE, RETEST

Table 30 presents the data for the Third Grade Retest group.¹ This gives the tests in rank order as measured by Criterion I for the group of 50 and by Criterion II for the group of 91. This table gives also the r , PE and $1-k$ for each test. $1-k$ is derived from the k value of the co-efficients of correlation as given in Miner's Tables, $\sqrt{1-r^2}$.

TABLE 30

THIRD GRADE RETEST—SECOND YEAR. RANKS OF TESTS BY
PARTIAL CORRELATION

50 CHILDREN, CRITERION I					91 CHILDREN, CRITERION II				
Test	Rank	r	PE	$1-k$	Test	Rank	r	PE	$1-k$
Otis ¹	1	.502	.071	.135	Myer	1	.868	.017	.503
Dn.	2	.478	.074	.122	Otis	2	.857	.019	.485
Hag.	3	.456	.075	.110	Dn.	3	.801	.025	.401
Kg.	4	.452	.076	.108	Kg.	4	.781	.028	.375
Myer	5	.374	.082	.073	Hag.	5	.711	.035	.297

¹ This table should be read: The Otis test ranks first with r .502, PE .071, $1-k$.135, etc. The measure $1-k$ indicates the improvement in estimate by a test over an estimate by chance, which in this test is .135.

The changes in rank shown by Criterion II as compared with the ranking by Criterion I are as follows:

Test	Criterion I	Criterion II	Change in rank
Otis	1	2	—1
Dearborn	2	3	—1
Haggerty	3	5	—2
Kingsbury	4	4	0
Myers	5	1	+4

One test, Kingsbury, does not change its rank; Otis and Dearborn drop down one step from first to second and from second to third respectively, Haggerty drops two steps and Myers rises four steps.

When considered on the basis of upper half (3 ranks) and lower half (2 ranks) Otis and Dearborn retain their positions in the upper half and Kingsbury retains its position in the lower half.

The advancement of Myers test to first place when measured by Criterion II indicates, of course, its greater degree of correspondence to the average of the five tests than to

¹ For statement regarding original data see page 32.

the IQ of the Stanford-Binet test. The reason why Myers test is so much more affected than the others is not at once evident. Myers IQ's ran very high in this grade and since Criterion II is the average of the GIQ's of the five tests it may be that this test had a disproportionate influence on Criterion II. This is another instance where the suggested study of another basis for Criterion II might help to solve the mystery.

The range of correlations with Criterion II is from r .870 to r .713 or .157 points, and with Criterion I is from r .502 to r .374 or .128 points. This is the first instance where the range of correlations has been found to be less about Criterion I than about Criterion II. This strengthens the feeling that Criterion II may not be the best measure as now composed.

When considered as to the advantage of using a test as shown by 1-k it is found that there is a range of advantage for correlation with Criterion II from about 51% to about 30%, a difference of 21%, and for correlation with Criterion I from about 14% to about 7%, a difference of 7%, but this difference shows double the efficiency of Otis test over Myers test as measured by 1-k of the correlations with Criterion I.

The ranks of the tests may be compared with those of the corresponding group for the first year. These rankings are shown in Table 31.

TABLE 31

RANKING OF TESTS OF THIRD GRADE AND OF THIRD GRADE RETEST
BY CRITERION I AND CRITERION II

Group of 50 children		Whole group	
First year	Second year	First year, 204	Second year, 91
1. Haggerty	1. Otis	1. Dearborn	1. Otis
2. Myers	2. Dearborn	2. Kingsbury	2. Myers
3. Otis	3. Haggerty	3. Myers	3. Dearborn
4. Kingsbury	4. Kingsbury	4. Otis	4. Kingsbury
5. Dearborn	5. Myers	5. Haggerty	5. Haggerty

A glance through these lists for identity of ranks shows:

Otis twice appears in first rank.

Myers twice appears in second rank.

Kingsbury three times appears in fourth rank.

Haggerty twice appears in fifth rank.

Divided on the basis of halves the result is as follows:

Half	Dearborn	Haggerty	Kingsbury	Myers	Otis
Upper	3	2	1	3	3
Lower	1	2	3	1	1

According to this rating Dearborn, Myers and Otis tests tie for first rank. Haggerty test takes fourth rank and Kingsbury fifth rank.

With only five tests smaller divisions are of less value than with a larger number of tests but for what it is worth the range is divided into three parts, high two, middle two and low one. This shows:

Part	Dearborn	Haggerty	Kingsbury	Myers	Otis
High	2	1	1	2	2
Middle	1	1	3	1	2
Low	1	2	0	1	0

According to this rating there is a little change in the former ranking.

Otis ranks first because it has no rank in the low part.

Dearborn and Myers tests tie for second rank.

Haggerty and Kingsbury follow but this time Kingsbury is above Haggerty with three ranks in the middle part and none in the low part, while Haggerty had one rank in the middle part and two ranks in the low part.

The two rankings compared show:

UPPER THREE, LOWER TWO		HIGH TWO, MIDDLE TWO, LOW ONE	
Test	Rank	Test	Rank
Otis	2	Otis	1
Dearborn	2	Dearborn	2½
Myers	2	Myers	2½
Haggerty	4	Kingsbury	4
Kingsbury	5	Haggerty	5

There are only two differences in the two rankings. Otis test is raised above the two with which it tied in the division into halves and Haggerty and Kingsbury test exchange places,

giving Kingsbury the advantage when the division is into three parts. Other things being equal the rating of tests in the smaller divisions would give greater accuracy in ranking.

In the first year's results Myers was the only test which was ranked in the upper three tests by both Criteria I and II and no test was placed by both criteria in the lower half. On this basis there is a confirmation of the Myers rating in the composite ranking. It is, however, worth noticing that the Myers test held second rank the first year when correlated with Criterion I and fifth rank the second year for the same group with Criterion I.

When the first year's results are divided into three parts the only agreement is in placing Otis test in the middle third. This does not agree with the composite ranking which gives it first place.

While considerable variation is shown in the results by different methods of rating the tests in these groups, the composite ratings showing the number of times each test appears in a particular part of the range are as a whole very closely in agreement whether the range is divided into halves or into thirds.

Since no tests were given in the fourth grade the first year, there is no comparable data for this second year work from the viewpoint of grade location for the first year. The third grade can be made a basis for comparison with lower grade work the second year. This comparison will be made later.

CHAPTER 6

CONCLUSIONS

THE FINDINGS OF THIS STUDY

Attention is called first to the fact that the group of children studied were, so far as chronological age is concerned, a satisfactory representation of the children in these grades in the white schools of Baltimore. The total population of these grades as shown in Table 2, Chapter 2, was 29,165. Whatever value would attach to such a representative group should be recognized here.

Second, there was a careful selection of schools to yield a fair representation as to racial and economic status of the children.

Third, all tests were given and scored by the same person, with the exception of one second grade class of 26 which had so recently been given the Pintner-Cunningham test by the Bureau of Research of Baltimore City that their scores were used for this group. The uniformity in giving and scoring would be expected to eliminate any errors which might arise from a difference in understanding of directions. To insure accuracy all scores used in this study were checked after the second year's study was completed. All calculations have had the most rigorous checking in the effort to eliminate chance errors.

Fourth, the use of the groups of 50 gave the possibility of using the Stanford-Binet test as the basis of one of the criteria. The IQ of this test was used as Criterion I. This is, as a rule, regarded as one of the standard criteria for evaluating tests.

Fifth, the criterion which was to include in some form the IQ's of the group tests themselves was established by a careful checking of different combinations of these GIQ's alone and in combination with the Stanford-Binet IQ.

In the first year the investigation was extended to include the ranking of the tests by their median deviation from

Criterion II which could be used in common for the whole group for each grade and for the group of 50 which was taken from the whole group. Also in the first year a study was made of the desirability of tests combined in pairs with every other test used in the grade. This proved to be a real addition to the study.

The use of Criterion II in the study of the work of both years gives the added advantage of gaining the confirmation or contradiction which comes from the results for the whole group as compared with the results for the groups of 50 when correlated with Criterion I.

In the summary of the study of the first year at the close of Chapter 4 each test was located as to its ranking in all grades where it was given, on the bases of the three criteria used in the first year.

Satisfactoriness of a test was determined by its frequency of location toward the upper end of the range of ranks rather than toward the lower end of the range. The distinction was made between the rankings by the median deviation and by correlation on the basis of the difference in meaning of these two. The ranking by median deviation indicates the closeness of correspondence of the median deviation of any test to the criterion. This is probably the best single measure of the accuracy with which individuals are measured by the test. If another measure is desired Table 5 in Chapter 3 gives also the Q and range of the deviations which offer a fair idea of the spread of these deviations.

The correlation of a test on the other hand takes account of relative position of individuals in the group instead of the individuals' relation to a particular point in the range of measures. For these reasons the two rankings are kept separate. If they agree it strengthens the indication of the general value or lack of value of the test. If they do not agree it may mean that the test is good for one form of measurement but not for the other.

In the use of correlation it should be held in mind that partial correlation is used in all these discussions so that chronological age is ruled out so far as its effect can be determined by the use of this method.

By the median deviation from Criterion II the tests were ranked high (H), middle (M), or low (L), for the grades where they were used as shown below. A blank space in the column indicates that the test was not given to the grade. Middle rank is given when in one group the test ranks in the upper half and in the other ranks in the lower half.

TABLE 32
RANKING OF TESTS FIRST YEAR BY MEDIAN DEVIATION FROM
CRITERION II

Test	Kindergarten	First	Second	Third grades
Cole-Vincent ¹	H	H		
Dearborn	M	H	H	H
Detroit		H		
Haggerty		L	L	L
Kingsbury		L	M	H
Myers		L	L	M
Otis		L	M	H
Pintner-Cunningham	L	M	H	
Pressey		M	H	

¹ This table should be read: Cole-Vincent test, when measured by the median deviation of its IQ's from Criterion II, ranks high in the kindergarten, and high in the first grade, etc.

From these results the suggested conclusions are, that, for measuring individual children in harmony with their measurement by the average of the tests given to their grade:

Cole-Vincent test ranks high for both kindergarten and first grade.

Dearborn ranks high for first, second and third grades and has middle rank for kindergarten.

Detroit ranks high for first grade.

Haggerty ranks low for first, second and third grades.

Kingsbury ranks low for first grade, middle for second, and high for third grade.

Myers ranks low for first and second grades and middle for the third grade.

Otis ranks low for first grade, middle for second and high for third.

Pintner-Cunningham ranks low for kindergarten, middle for first grade and high for second.

Pressey ranks middle for first grade and high for second.

When listed as ranked by correlation these tests stand as shown in Table 33. The plan is used of ranking the test high if placed in the upper half by both criteria, low if placed in the lower half by both criteria and middle if placed in the upper half by one and in the lower half by the other criterion. A blank indicates that the test was not given.

TABLE 33
RANKING OF TESTS FIRST YEAR BY PARTIAL CORRELATION.
CRITERIA I, II

Test	Kindergarten	First	Second	Third grades
Cole-Vincent ¹	M	M		
Dearborn	L	L	L	M
Detroit		M		
Haggerty		L	L	M
Kingsbury		M	H	M
Myers		L	H	H
Otis		H	H	M
Pintner-Cunningham	H	H	M	
Pressey		M	M	

¹ This table should be read: Cole-Vincent test, when ranked by partial correlation with the two criteria, ranks middle for both kindergarten and first grade, etc.

This rating places the single tests as follows:

Cole-Vincent in middle rank for both kindergarten and first grade.

Dearborn ranks low for kindergarten, first, second and third grades.

Detroit ranks middle for first grade.

Haggerty ranks low for first and second grades and middle for third grade.

Kingsbury ranks middle for first grade, high for second, and middle for third grade.

Myers ranks low for first and high for second and third grades.

Otis ranks high for first and second grades and middle for third.

Pintner-Cunningham ranks high for kindergarten and first grade and middle for second grade.

Pressey ranks middle for first grade and for second grade.

The ranking of these tests as shown by the frequency of the satisfactoriness of their combination with other tests is shown in Table 34. Low means low in the range of ranks, high means high in the range of ranks, and middle, in the middle of the range of ranks.

TABLE 34
RANKING OF TESTS FIRST YEAR AS SHOWN BY THEIR GROUPING IN
PAIRS AND CORRELATED WITH CRITERION I

Test	Kindergarten	First	Second	Third grades
Cole-Vincent ¹	L	L		
Dearborn	M	L	L	M
Detroit		M		
Haggerty		L	L	H
Kingsbury		L	H	L
Myers		H	H	H
Otis		H	M	H
Pintner-Cunningham	H	H	M	
Pressey		H	M	

¹ This table should be read: Cole-Vincent test, when rated by the satisfactoriness of its combination with other tests as shown by correlation by Criterion I, ranks low for both kindergarten and first grade, etc.

For purposes of combining with other tests this rating places the tests as follows:

Cole-Vincent ranks low in both kindergarten and first grade.

Dearborn ranks middle in kindergarten, and third grade, and low in first and second grades.

Detroit ranks middle in first grade.

Haggerty ranks low in first and second grades and high in third grade.

Kingsbury ranks low in first and third grades and high in second grade.

Myers ranks high in first and second and third grades.

Otis ranks high in first and third grades and middle in second grade.

Pintner-Cunningham ranks high in kindergarten and first grade and middle in second grade.

Pressey ranks high in first and middle in second grade.

A caution should be held in mind with regard to this last list of ratings, that it is apparently of considerable consequence with which other tests each test is paired. Therefore, it is not safe to take a test which ranks high in this list and pair it indiscriminately with any other test.

The confirmation of the ratings of these tests has been shown in Chapter 5 by making a composite ranking, by showing the frequencies in a particular part of the range when measured by the two Criteria I and II, for both first and second years. Here it was found profitable to divide the range roughly into thirds for the rankings by Criterion I and Criterion II the first year and by Criterion I and Criterion II the second year. This gives a possible frequency of any test of four. If the mode of the ranks is in the upper third the rating is high (H), if in the lower third the rating is low (L), if in the middle third the rating is M+ if the tendency is to the upper third and M- if the tendency is toward the lower third.

TABLE 35

RANKING OF TESTS BY A COMPOSITE OF THE RANKINGS BY CRITERION I AND CRITERION II FOR THE FIRST AND SECOND YEARS

Test	Kindergarten	First	Second	Third grades
Cole-Vincent ¹	M	M—		
Dearborn	L	M+	M	M+
Detroit		L		
Haggerty		L	L	M—
Kingsbury		M	H	M+
Myers		M—	H	M+
Otis		H	M+	H
Pintner-Cunningham	H	H	H	
Pressey		M+	M—	

¹ This table should be read: Cole-Vincent test, when ranked by the composite ranking stated at the top of this table, ranks middle in kindergarten and below middle in the first grade.

By this composite rating:

Cole-Vincent ranks middle for kindergarten and below middle for first grade.

Dearborn ranks low for kindergarten, above middle for first and third grades and middle for second grade.

Detroit ranks low for first grade.

Haggerty ranks low for first and second and below middle for third grade.

Kingsbury ranks middle for first grade, high for second and above middle for third grade.

Myers ranks below middle for first grade, high for second grade and above middle for third.

Otis ranks high for first and third grades and above middle for second grade.

Pintner-Cunningham ranks high for kindergarten, first and second grades.

Pressey ranks above middle for first grade and below middle for second grade.

Certain of these rankings by use of the composite rank may be unjust because of the fact that the class as a whole was advanced a year at the second testing and this automatically led to a disadvantage for particular tests which were used in grades for which they were not designed. This is true of the Cole-Vincent test in the first grade retest where the composite ranking shows less than middle in rank; with the Detroit, which is planned only for first grade and, when used for first grade retest, seemed out of place as related to its satisfactory position the first year. It is also true of the Pintner-Cunningham and Pressey tests the second year with the second grade retest group. There is no disadvantage evident for the Pintner-Cunningham second grade retest as it stands high in the composite. There may be some for the Pressey test, for the same group.

Another comparison may be made on these tests in the effort to gain further light. This is the comparison of the results of each retest group with the results of the group one grade in advance of them the first year. This will give a comparison of the kindergarten retest group with the first grade (first year) group; of the first grade retest group with the second grade (first year) group; and of the second grade retest group with the third grade (first year) group. Because no fourth grade group was tested the first year the third

grade retest group cannot be compared with any equivalent from the first year. Only tests given to both groups can be considered.

TABLE 36
RANKING OF TESTS BY CRITERION I AND CRITERION II FOR
"FIRST GRADE CHILDREN"

First grade, first year		Kindergarten, retest, second year	
Criterion I	Criterion II	Criterion I	Criterion II
1. Pintner-Cunningham ¹	1. Pintner-Cunningham	1. Dearborn	1. Pintner-Cunningham
2. Cole-Vincent	2. Dearborn	2. Cole-Vincent	2. Dearborn
3. Dearborn	3. Cole-Vincent	3. Pintner-Cunningham	3. Cole-Vincent

¹ This table is to be read: For the first grade the first year by Criterion I and Criterion II Pintner-Cunningham test ranks highest; for the kindergarten retest the second year by Criterion I Dearborn test ranks highest; and by Criterion II Pintner-Cunningham test ranks highest, etc.

The following summary may be made of these ranks as to frequency:

Part	Cole-Vincent	Dearborn	Pintner-Cunningham
High	0	1	3
Middle	2	2	0
Low	2	1	1

This shows Pintner-Cunningham ranking first; Dearborn ranking second; and Cole-Vincent ranking third.

TABLE 37
RANKING OF TESTS BY CRITERION I AND CRITERION II FOR
"SECOND GRADE CHILDREN"

Second grade, first year		First grade, retest, second year	
Criterion I	Criterion II	Criterion I	Criterion II
1. Kingsbury	1. Otis	1. Dearborn	1. Kingsbury
2. Pintner-Cunningham	2. Kingsbury	2. Pintner-Cunningham	2. Pressey
3. Myers	3. Myers	3. Pressey	3. Otis
4. Otis	4. Pressey	4. Kingsbury	4. Pintner-Cunningham
5. Dearborn	5. Pintner-Cunningham	5. Otis	5. Dearborn
6. Pressey	6. Dearborn	6. Myers	6. Myers
7. Haggerty	7. Haggerty	7. Haggerty	7. Haggerty

This gives the following frequency of rankings in high three, middle two and low two.

Part	Dn.	Hag.	Kg.	Myer	Otis	P-C	Pr.
High	1	0	3	2	2	2	2
Middle	2	0	1	0	2	2	1
Low	1	4	0	2	0	0	1

By this ranking:

Kingsbury ranks first.

Otis and Pintner-Cunningham tie for second place.

Pressey ranks fourth.

Myers and Dearborn dispute the fifth rank. Myers has two against Dearborn's one in the high rank but has none in the middle rank, so that if the ranking is on the basis of frequency above the low rank Dearborn takes fifth rank and Myers sixth rank.

Haggerty has undisputed claim to seventh rank.

TABLE 38
RANKING OF TESTS BY CRITERION I AND CRITERION II FOR
"THIRD GRADE CHILDREN"

Third grade, first year		Second grade, retest, second year	
Criterion I	Criterion II	Criterion I	Criterion II
1. Haggerty	1. Dearborn	1. Haggerty	1. Kingsbury
2. Myers	2. Kingsbury	2. Otis	2. Dearborn
3. Otis	3. Myers	3. Myers	3. Myers
4. Kingsbury	4. Otis	4. Dearborn	4. Haggerty
5. Dearborn	5. Haggerty	5. Kingsbury	5. Otis

Dividing these lists into high two, middle two and low one, the following frequencies result:

Part	Dearborn	Haggerty	Kingsbury	Myers	Otis
High	2	2	2	1	1
Middle	1	1	1	3	2
Low	1	1	1	0	1

By this grouping it is shown that Dearborn, Haggerty and Kingsbury hold exactly the same positions so far as distribution of their ranks is concerned. Myers comes fourth with one rank in the high ranks and three in middle rank. It is possible that Myers might dispute the rank with the other three because of having all its ranks above the lower third. The writer is inclined to place more value on the two ranks in the high division.

Otis here ranks fifth for third grade use.

It will be recalled that two previous studies had been published which were made for the purpose of comparing some of these group tests of intelligence for primary grades. It was stated in reporting these studies that no direct com-

parison of the studies could be made because of the difference in criteria used, and because of the difference in the unit of measurement. There will, however, be some interest in comparing the rankings found in the earlier studies with the rankings found in the present study. Of course this can only be done for the tests which are included in at least one of the earlier studies. These data are arranged in Tables 39, 40 and 41.

TABLE 39

(Miss Streitz's Study)

CRITERION SIMPLE CORRELATION WITH TEACHERS' ESTIMATES AVERAGE OF GRADES 1 AND 2					
Single tests		Single tests when combined in pairs			
	Rank	Methods:	Weighting	Average of scores	Total
Dearborn	1.	Upper	Dearborn	2	4
Haggerty	2.		Myers	2	4
Pressey	3.		Pressey	3	5
Myers	4.		Haggerty	1	1
		Lower	Dearborn	1	2
			Myers	1	2
			Pressey	0	1
			Haggerty	2	5

TABLE 40

(Miss Velie's Study)

CRITERION HIGHEST CORRELATION OF TEST WITH TEACHER'S ESTIMATE

Grade 1		Grade 2	
Test	Rank	Test	Rank
Pintner-Cunningham	1	Dearborn	1
Detroit	2	Pintner-Cunningham	2
Cole-Vincent	3	Cole-Vincent	3
Dearborn	4	Detroit	4

TABLE 41

PRESENT STUDY. FIRST YEAR
CRITERIA I AND II

First grade		Second grade	
Test	Rank	Test	Rank
Pintner-Cunningham	H	Myers	H
Cole-Vincent	M	Pintner-Cunningham	M
Detroit	M	Pressey	M
Pressey	M		
Dearborn	L	Dearborn	L
Myers	L	Haggerty	L
Haggerty	L		

The comparison of these tables is unsatisfactory regardless of the results found because of the lack of similarity of the data considered. If the comparison yields anything it will be merely the demonstration of the similarity or the difference of rankings of the tests by different methods of studying them. It will be found that Pintner-Cunningham test is ranked high in the two studies where used; that Pressey ranks high in the two studies but the test used in Miss Streitz's study was an earlier edition than the one used in the present study; that Dearborn on the three studies gives irregular ranking but is divided about equally between upper and lower halves. The Dearborn test is the only test which was used in the three studies.

Haggerty holds a low rank in both studies where it is used, except in the combination of the two grades where it is correlated alone with teachers' estimates, where it ranks second of the four tests given.

Myers test stands a little above the middle rank when its ranks in the two studies are compared.

Detroit stands twice high and once low and Cole-Vincent stands high once and low twice.

The Pintner-Cunningham and Pressey tests are the ones which, when used in these studies, stand out as satisfactory. Each of them was used in two of the studies. The Otis and Kingsbury tests which several times showed up to good advantage in the present study were not used in either of the earlier studies. A use of these tests in either of the other studies might conceivably have changed the ranks of the tests used.

There have been presented the evaluations of these tests which have come out of this study together with many comparisons of them. So far as the selection of children and the methods of giving the tests and handling the data are accepted as valid these ranks have some importance. It is not assumed that they answer all questions or will be uniformly valid in all circumstances.

TEST SUGGESTIONS TO SCHOOL OFFICERS

If tests are to be of value in school situations it is because they provide a means of answering questions which arise in the organization and administration of the school. Besides departments of research or where these are lacking, superintendents, supervisors, heads of schools, or classroom teachers may have problems for solutions for which they wish to use standardized tests. Some basis for determining before the work is undertaken the probable degree of satisfactoriness of the use of particular tests would be of considerable value to those wishing to use them.

The present study offers an answer to the question of the relative desirability of the tests studied in relation to three different problems which may arise.¹

First, educators may wish to use a group test which will give a measurement of individual children as near as possible to the measurement which would be obtained by giving all the group tests which were available for the study. The study shows the following answer to the question as to which of the tests used meet this need.

For the kindergarten: Cole-Vincent test is the best test as measured by the median, quartile, and range of the deviations of the individual scores from the average of the individual scores on the three tests given. Dearborn is the next in rank.

For the first grade: Dearborn and Cole-Vincent rank highest of the nine tests given, with the Detroit test a close third.

For second grade: The tests which rank highest of the seven used are the Pressey, Dearborn, and Pintner-Cunningham. There is a slight difference in the order for the group

¹ The statements here given are on the basis of the first year's study: 1. Because there were twice as many children included; 2. Because some of the studies were only made the first year; 3. Because it was found that the second year's results did not make much change in those of the first year, and where there was change it might in some cases be attributed to inadequacy of the test to the second year classification of the children. (This was explained in the discussion.) Those interested should when possible refer to the study in detail.

of 50 and for the group of 235 but in each group these three are the three highest.

For third grade: Otis, Dearborn, and Kingsbury tests come in the first three ranks for the group of 204 and in the first three and a half ranks for the group of 50. This slight difference is caused by the fact that the Myers test appears in second rank in the group of 50 and that Dearborn and Kingsbury have in this group the same median deviation.

It appears that if what is wanted is a single group test which will give measurements of individuals closely comparable to those given by the average of several group tests for these individuals the Cole-Vincent test is acceptable for both kindergarten and first grade; the Dearborn test is acceptable for each of the four grades; the Detroit test is acceptable for first grade; the Pressey and Pintner-Cunningham tests for second grade; and the Otis and Kingsbury tests for third grade.

Second, the problem for solution may be that of ranking the children. In this case the exact measurement of each individual child may be a matter of secondary importance, the chief problem being to rank the children as to degree of intelligence so as to place them in relatively homogeneous groups. This might give an X, Y, Z, grouping. For this purpose there are two standards which can be derived from the present study. The desire may be to rank the children by a single group test as near as possible to what their ranking would be if they each were given the Stanford-Binet test. The present study shows that certain tests give greater satisfaction in this matter than others.

For kindergarten: Pintner-Cunningham is best and Cole-Vincent comes next.

For first grade: Detroit comes first, then Pintner-Cunningham.

For second grade: The best test is Kingsbury. This is followed by two tests almost identical in rating—Pintner-Cunningham and Myers.

For third grade: Haggerty comes first and Myers is second.

If a ranking similar to that given by the Stanford-Binet test is desired this study indicates that Pintner-Cunningham test is best for the kindergarten. It also shows good reliability for first and second grades. Kingsbury test is best for second grade and Haggerty is best for third grade.

Third, the other standard for determining rank which may be used is the comparison of the ranking of a given test with that which would be found by the use of the average of the whole number of tests given to the grade in this study. The results on this basis are as follows:

For kindergarten: Pintner-Cunningham holds first place, Cole-Vincent has second place and Dearborn is last. While Pintner-Cunningham is the best there is so little difference in the correlations that no serious deviation in ranking would occur in using any of these tests, on the basis of this criterion. Of course it will be recognized that there is naturally a closer relation to the criterion of which a factor is itself one-third than when any one factor is only one-fifth, one-seventh, or one-ninth, as in the other grades.

For first grade: Otis, Pintner-Cunningham, and Pressey are the three highest of the nine tests used. There is so little difference among them that any one of these might be accepted to meet this standard.

For second grade: Otis, Kingsbury, and Myers are the three highest.

For third grade: Dearborn ranks first by a considerable margin.

If ranking similar to what would come from use of the average of all the tests is desired we find the following situation: Pintner-Cunningham stands first in kindergarten and second in first grade. It has a correlation above .80 in the second grade. Otis stands first in both first and second grades. Pressey stands third in first grade and fourth in second grade. Dearborn which stands first in third grade also has a correlation above .80 in second grade. Kingsbury and Myers stand second and third in rank respectively in both second and third grades.

If it is desired to use a second test which will increase as much as possible the accuracy of the measurement, the results of this study indicate that for the kindergarten the Dearborn test should be added to the Pintner-Cunningham; for the first grade the Pressey test will improve any test with which it is combined; for the second grade the Kingsbury test will improve any test with which it is used; for the third grade Myers test combined with Haggerty gives the best combination. Otis combined with Haggerty comes very close to this, and the Myers and Otis combination is only slightly lower. Tables showing the combinations of the tests should be consulted for the specific values of particular combinations.

ADDITIONAL PROBLEMS

Many questions have arisen in the handling of these data. There has been an effort to disregard all side issues and attend for the time only to the results which have accumulated on the basis of the criteria accepted for the study. One question which has sought recognition is that as to whether some other means of choosing the criterion called Criterion II might have led to greater fairness in measuring the tests. It is the intention of the writer to attack this problem and if it seems desirable to carry out different lines of investigation in an effort to find an answer to this question.

Another question to which the accumulated data should produce an answer is as to the relation of teachers' estimates to test ratings. If the whole 800 tests of the first year were used it seems probable that the results would yield a reasonable answer to this problem.

Another interesting opening for study is to investigate the results for the 200 children who had all tests for both years and also the Stanford-Binet test, to discover individual growth from year to year.

The matter of practice effect of testing has not been touched because there seemed to be no need for it in the direct path toward the solution of the problem of this study.

An analysis of the results for each test might be made for the purpose of discovering if possible what test items

yield the best results. If such a study could be carried out to a successful issue it should yield data for the formulation of a super-test which would give better results than any test now available.

The question of the direction of the deviations of a test from the criterion was just touched in an exploratory study and seemed to offer interesting possibilities.

A study which would devise a scheme for the best weighting of these tests for combination, probably using the Multiple Regression Equation as the determinant of weighting values, might throw considerable light on the usefulness of these tests.

Such are some of the questions which have arisen during the course of this study and have been neglected for the main issue. These seem to indicate that the basic data of this study have not been drained dry and also that there is an open field for others to conduct investigations in the same or neighboring fields which will be rewarding.

BIBLIOGRAPHY

TESTS AND MANUALS

(The tests which are starred were used in this study.)

- BALDWIN, B. T., and STECKER, L. E. *The Psychology of the Pre-School Child*. New York, D. Appleton and Company, 1924.
- *COLE-VINCENT TEST FOR SCHOOL ENTRANTS. Test and Manual. Emporia, Kansas. Kansas State Teachers College.
- *DEARBORN, W. F. *The Dearborn Group Tests of Intelligence*. Series I and Manual, Philadelphia, Pa., J. B. Lippincott and Company.
- *ENGLE, A. M. *Detroit First Grade Intelligence Test and Manual*, and *Directions for Use*. Yonkers-on-Hudson, N. Y., World Book Company.
- GAMBRILL, B. L., and FARWELL, L. *An Analytical List of Kindergarten-Primary Tests of Intelligence and Achievement*. New Haven, Conn., Whitlock's Book Store, Inc.
- GESELL, A. *The Mental Growth of the Pre-School Child*. New York, Macmillan Company, 1925.
- *HAGGERTY, M. E. *Haggerty Intelligence Examination, Delta I and Manual*. Yonkers-on-Hudson, N. Y., World Book Company.
- JOHNSON, B. J. *Mental Growth of Children in Relation to the Rate of Growth in Bodily Development*. New York, E. P. Dutton and Company, 1925.
- *KINGSBURY, F. A. *Primary Group Intelligence Scale, and Manual*. Bloomington, Ill., Public School Publishing Company.
- LOWELL, F. A. *A Group Intelligence Test for Primary Grades*. *Journal of Applied Psychology*, Vol. 3 (1919) pp. 199-214.
- MERRILL-PALMER PERFORMANCE TESTS. *Merrill-Palmer School*, Detroit, Michigan.
- *MYERS, G. C. *Myers Mental Measurement Test and Manual*. Carlisle, Pa., The Sentinel.
- *OTIS, A. S. *Group Intelligence Scale, Primary Examination and Manual*. Yonkers-on-Hudson, New York, World Book Company.
- *PINTNER, R., and CUNNINGHAM, B. V. *Primary Mental Test and Manual*. Yonkers-on-Hudson, N. Y., World Book Company.
- *PRESSEY, L. W. *Primary Classification Scale and Manual*. Bloomington, Ill., Public School Publishing Company.
- STUTSMAN, RACHEL. *Performance Tests for Children of Pre-School Age*. *Genetic Psychology Monograph I*, No. 1 (1927), pp. 3-65.
- TERMAN, L. M. *The Measurement of Intelligence*. Boston, Houghton, Mifflin Company, 1916.

DISCUSSION OF TESTS AND REPORTS OF THEIR DEVELOPMENT AND USE

- ARTHUR, GRACE. "An Application of Intelligence to the Problem of School Retardation," *School and Society*, Vol. 10 (1919) pp. 614 ff.
- ARTHUR, GRACE. "Eliminating First Grade Failures Thru Control of Intellectual, Physical and Emotional Factors." *School and Society*, Vol. 15 (1922) pp. 474-484.

- BERRY, CHARLES S. "Classification of Tests of Intelligence of 10,000 First Grade Pupils." *Journal of Educational Research*, Vol. 6 (1922), pp. 185-203.
- BINZELL, ALMA L. "General Characteristics and Work of Measurements in Grades and Need of These in the Kindergarten Field." *Kindergarten and First Grade*, Vol. 3 (1918), pp. 353-357.
- BISHOP, O. "What Is Measured by Intelligence Tests?" *Journal of Educational Research*, Vol. 9 (1924), pp. 29-38.
- CHASSELL, C. F. and L. M. "Survey of First Three Grades of Horace Mann School." *Journal of Educational Psychology*, Vol. 12 (1921), pp. 77-81, 243-252.
- CUNEO, I., and TERMAN, L. M. "Stanford-Binet Tests of 112 Kindergarten Children and 77 Repeated Tests." *Pedagogical Seminary*, Vol. 25 (1918), pp. 414-428.
- CUNNINGHAM, B. V. The Prognostic Value of a Primary Group Test. *Columbia University Contributions to Education*, No. 139, 1923.
- DICKSON, V. E. Mental Testing of all Kindergarten Children. *National Education Association Proceedings*, 1920, pp. 196-198.
- DICKSON, V. E. "What First Grade Children Can Do in School as Related to What is Shown in Mental Tests." *Journal of Educational Research*, Vol. 2 (1920), pp. 475-480.
- DICKSON, V. E. *Mental Tests and the Classroom Teacher*. Yonkers-on-Hudson, N. Y., World Book Company, 1923.
- FREEMAN, FRANK N. *Mental Tests*. Boston, Houghton, Mifflin Company, 1925.
- GOODENOUGH, F. L. *Measurement of Intelligence by Drawings*. Yonkers-on-Hudson, N. Y., World Book Company, 1926.
- GRAVES, K. B. Influence of Specialized Training on Tests of General Intelligence. *New York, Columbia University Contributions to Education*, No. 143, 1924.
- HENMON, V. A. C., and STREITZ, RUTH. "A Comparative Study of Four Group Scales for the Primary Grades." *Journal of Educational Research*, Vol. 5 (1922), pp. 185-194.
- HICKS, V. C. "The Value of Mental Age Tests for First Grade Entrants." *Journal of Educational Psychology*, Vol. 6 (1915), pp. 157-166.
- HOLLEY, C. E. *Mental Tests for School Use*. Urbana, Illinois, University of Illinois, 1920.
- JOHNSON, O. J. "Group Intelligence Examinations for Primary Children." *Journal of Applied Psychology*, Vol. 6 (1922), pp. 403-416.
- LOWELL, F. A. "Preliminary Report of Some Group Tests of General Intelligence." *Journal of Educational Psychology*, Vol. 10 (1919), pp. 323-344.
- MCCALL, W. S. *How to Measure in Education*. Boston, Houghton, Mifflin Company, 1923.
- MYERS, G. C. "Teachers vs. Mental Tests as Prophets of School Progress." *School and Society*, Vol. 16 (1922), pp. 300-303.
- MYERS, C. E. and G. C. "A Group Intelligence Test." *School and Society*, Vol. 10 (1919), pp. 355-360.
- PINTNER, R., and CUNNINGHAM, B. V. "The Problem of Group Intelligence Tests for Very Young Children." *Journal of Educational Psychology*, Vol. 13 (1922), pp. 465-472.
- PINTNER, R., and CUNNINGHAM, B. V. "The Pintner-Cunningham Primary Test." *Journal of Educational Psychology*, Vol. 18 (1927), pp. 52-58.

- PINTNER, R. "Intelligence Testing." New York, Henry Holt and Company, 1923.
- PRESSEY, S. L. and L. W. "Group Point Scale for Measuring General Intelligence with First Results from 1100 School Children." *Journal of Applied Psychology*, Vol. 2 (1918), pp. 250-269.
- PRESSEY, L. W. "Group Scale of Intelligence for Use in First Three Grades." *Journal of Educational Psychology*, Vol. 10 (1919), pp. 297-308.
- PRESSEY, L. W. "The Primary Classification Tests." *Journal of Educational Research*, Vol. 9 (1924), pp. 305-314.
- PRESSEY, S. L. and L. W. "Measuring the Usefulness of Tests in Solving School Problems." *School and Society*, Vol. 12 (1920), pp. 531-534.
- STREITZ, RUTH. "An Experimental Investigation of Four Intelligence Scales for Primary Children." Master's Thesis, September 1922, University of Chicago.
- TERMAN, L. M. "The Intelligence of School Children." Boston, Houghton, Mifflin Company, 1919.
- TERMAN, L. M. and OTHERS. *Intelligence Tests and School Reorganization*. Yonkers-on-Hudson, N. Y., World Book Co., 1923.
- TOWN, CLARA H. "A Mass Test for Use with Kindergarten and First Grade Children." *Journal of Applied Psychology*, Vol. 6 (1922), pp. 89-112.
- Twenty-First Yearbook, National Society for the Study of Education, Part II.
- VIELE, ADA B. "A Study of Four Primary Mental Tests." *Elementary School Journal*, Vol. 25 (1925), pp. 675-681.

VITA

Mary Lorette Dougherty, the daughter of the Reverend James Gregory, and Sarah Perry Dougherty, was born in Wyandotte, Kansas. The degree of Ph.B. in education was received from the University of Chicago in September, 1916, with both general and departmental honors. She was elected to Phi Beta Kappa (Beta of Illinois), in August, 1916. The degree of M. A. in education was conferred by the University of Chicago in September, 1917. In April, 1922, she was elected to Pi Lambda Theta (Lambda Chapter), the National Honor Fraternity for graduate women in education. Graduate study was carried on at the University of Chicago during the summer of 1924 and at the Johns Hopkins University during the school years 1923 to 1928. In February, 1928, the degree of Doctor of Philosophy was conferred by the Johns Hopkins University.

She taught in the primary schools in Kansas City, Kansas, from September, 1895, to June, 1915. She held the position of Assistant Professor of Education at the Iowa State Teachers College, Cedar Falls, Iowa, during the school year 1917-1918, and the next year the position of Professor of Education. In September, 1919, she went to the State Teachers College at Mankato, Minnesota, where she was for two years principal of the primary grades and instructor in primary education. During the years 1922 and 1923 she conducted the psycho-physical and mental tests in the university elementary school of the University of Chicago and directed the testing in the public schools of North Evanston, Illinois. Since September, 1923, she has been engaged at the Johns Hopkins University in teaching three classes and conducting the Educational Clinic. The three summers, 1925, 1926, 1927, she taught in the summer session at the University of Chicago.

Her published writings are as follows:

- Report on the Binet-Simon Tests Given to Four Hundred and Eighty-Three Children in the Public Schools of Kansas City, Kansas.* Jour. Educ. Psychol. 1914, 4, 338-352.
- The History of the Teaching of Handwriting in America.* The Ele. Sch. Jour., 1917, 18, 280-286.
- Results of the Use of an Experimental Course in Handwriting.* Master's Thesis, University of Chicago, August, 1917.
- An Experiment in Teaching Arithmetic in the Third Grade.* Ele. Sch. Jour. 1922, 22, 665-676.
- How to Teach Handwriting.* (Frank N. Freeman and Mary L. Dougherty) Boston, Houghton Mifflin Company, 1923.
- How to Teach Phonics.* Boston, Houghton Mifflin Company, 1923.
- The Relation Between Physical and Mental Development.* Ele. Sch. Jour. 1923, 24, 130-134.
- The Teaching of Phonics in the Primary Grades.* The Ele. Eng. Rev. 1924, 1, 181-183.
- Phonics in the Intermediate Grades.* Baltimore Bulletin of Education, Special Bul. on Reading, 1926.
- Phonics.* In the Human Interest Library, Chicago. The Midland Press, 1927.
- Learning to Write.* Normal Instructor and Primary Plans. 1926-1927. (5 articles.)

